Organizational Design and Individual Behavior

Inaugural-Dissertation

an der

Ludwig-Maximilians-Universität München

von

Benjamin Häusinger

Organizational Design and Individual Behavior

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2019
vorgelegt von

Benjamin Tobias Häusinger

Referent:	Prof. Dr. Klaus M. Schmidt
Korreferent:	Prof. Dr. Florian Englmaier
Promotionsabschlussberatung:	05.02.2020

Tag der mündlichen Prüfung: 20.01.2020

Namen der Berichterstatter: Klaus M. Schmidt, Florian Englmaier, Matthias Lang

Pentru Valentina

Acknowledgements

In the preparation of this dissertation my first supervisor and mentor, Klaus M. Schmidt, played the most influential and important role. His precise, thoughtful and therefore invaluable comments improved my research, my reasoning and this dissertation tremendously. I would like to thank him for his thoroughness, reliability and accessibility, for the frequent and fruitful discussions about my research, and in particular for his honest opinion and open words.

I would also like to thank Florian Englmaier, my second supervisor, for his excellent support over the last four years, for his belief in my abilities, for his constant encouragement and for his comments on my research. Moreover I thank Matthias Lang, who also serves as my third advisor, and Yves Le Yaouanq for many meetings and discussions.

Without support from the administrative side this dissertation would have been impossible; I thank Tamilla Benkelberg, Julia Zimmermann and Sabine Wilhelm-Kauf as well as our research assistants for their help, support and joyful conversations. I thank the Elite Network Bavaria and the German Science Foundation for their financial support via the doctorate program "Evidence-Based Economics" and the Collaborative Research Center Transregio "Rationality and Competition", respectively. During the PhD, I visited Kellogg School of Management and I thank Niko Matouschek for making this visit possible.

Over the course of the PhD, I was lucky to find many new friends and colleagues at the Seminar of Economic Theory, the Munich Graduate School of Economics and the Faculty of Economics. On various occasions they inspired, helped and advised me and, most importantly, made my life as a PhD student a wonderful experience. There are many people to thank, for all the discussions over coffee, lunches, dinners, beers and wine, in particular Hungni Chen, Tobias Rossmann, Christoph Winter, Matthias Wilhelm, Felix Peterhammer and Navid Sabet. I am especially grateful to Marvin Deversi who I met on day one of our PhD and who became a great colleague, outstanding economist and, most importantly, a very close friend.

The last four years have been an intellectual but also personal challenge in many respects. I thank everyone, including the above mentioned, who has helped me to write this dissertation and grow as a person. In particular I would like to thank my longtime friends, most notably Frieder Mitsch, Nicolai Fischer and Manuel Bühler, for all the days and nights full of conversations, discussions, and jokes, and for not taking themselves, let alone me, too seriously. Likewise, I thank Andreas Onken, Kerstin Caputo, Xinping Luo, Aleksandra Tomic and Leon Plahuta as well as Steve Krakosky for their continuing friendship.

Și mulțumesc frumos, Valentina. I am grateful to you beyond imagination. Thank you for your empathy, love and encouragement throughout the making of this dissertation; thank you for your humor and intellect that I am lucky to enjoy everyday. Thank you for making me smile and thinking the way I think. Especially, I thank you for letting me always letting you win in *table* (backgammon). Further I would like to thank Elena and Nicolae Voican as well as Andreea, Andrei and Matei Cretu Voican for their warmth, openness and hospitality.

Finally, I want to thank above all my parents Charlotte and Steven and my sisters Stephanie and Corinna. I would not be the person I am today without their unconditional love and support. Among countless other blessings, I am grateful for the optimism, confidence and peacefulness they have instilled in me and the role models they are to me. I am unable to express in words the weight of my gratitude and love. *Thank you.*

Benjamin Häusinger
Munich, September 2019

Contents

List of Figures

List of Tables

Introduction

Organizations constitute groups of individuals formed to govern and guide individual behavior; they are "a means of achieving the benefits from collective action" (Arrow 1974, p.33). That definition does not only include firms, but also public organizations, clubs, churches, and indeed any organized group of people (Bernard 1938). Within those, individuals interact in a variety of ways and are therefore strongly intertwined. They communicate, gather, and exchange information; they collaborate, help each other, and work together on projects; they compete for recognition, promotions, and bonuses.

It is these multi-faceted interactions that make the optimal structure of organizations, the *organizational design*, a matter of vast complexity. A firm must consider how to efficiently motivate employees, induce cooperative behavior, structure the decision-making process, enhance the flow of knowledge, and much more. For that reason, the optimal "design of organizations has become an object of inquiry" (Arrow 1964, p.398), and despite decades of research numerous questions remain unanswered (Gibbons and Roberts 2012).

This dissertation contains three chapters that help to understand the optimal design of organizations and the behavior of individuals therein. It provides insights and answers to (1) when and why individuals seek advice, (2) how the design of rules affects rule adherence, and (3) how organizations optimally structure promotions and managerial decision-making. The general approach of the dissertation is similar across all three chapters. In each chapter, I use theoretical models to investigate individual behavior and the impact of organizational design thereon. I rely extensively on tools and models developed in microeconomic theory and game theory. For instance, I analyze individual decision-making in Chapter 1 and Principal-Agent models in Chapters 2 and 3. Chapter 2 also includes results from an economic laboratory experiment.

Furthermore, the first and the second chapter analyze the decision problem of an individual with reputation concerns and thus build on a large literature on signaling games, starting with Spence (1973). I follow the approach by, e.g., Akerlof (1980), Bernheim (1994) and Bénabou and Tirole (2006); in both chapters an individual derives additional utility from her reputation or social image, i.e. when appearing smart (Chapter 1) or adherent to rules (Chapter 2). Two interpretations are consistent with that approach. The first one builds on a game theoretical argument and explains reputation utility as a simplification of a repeated game. In that case, higher reputation increases utility because it implies higher benefits in later periods. This interpretation is in line with the model in Chapter 1. There employees want to impress their supervisor, for example because a good reputation increases the chances of getting promoted which implies higher future wages. The second interpretation, based on findings in psychology and behavioral economics, explains utility from reputation as an intrinsic preference (Bénabou and Tirole 2006). Individuals feel better when others hold them in high regard. The model in the second chapter follows that idea; there individuals want to be perceived as rule-following, for instance because it is a social norm to adhere to rules (Sunstein 1996).

All three chapters in this dissertation are single-authored and self-contained, with separate appendices and one common bibliography at the end of the dissertation. This introduction proceeds with an overview over the three chapters to outline their respective motivation, intuition and main insights.

Overview

Chapter 1 relates to questions on employee cooperation and the flow of knowledge and information in organizations. It investigates the impact of reputation concerns and ability on an employee's incentives to seek advice. Advice and help are crucial for informed decision-making and efficient work; therefore organizations put special emphasis on encouraging advice-seeking when it is needed.

The chapter is motivated by an obstacle to advice-seeking, namely the common intuition that "by seeking help [i.e. advice], one publicly acknowledges incompetence" (Lee 2002, p.19). Following that idea, asking for advice reveals that the advice-seeker lacks the skills to solve a problem alone. Therefore individuals fear for their reputation and may not ask for advice or help when they need it. An inefficient equilibrium arises because of reputation concerns (Levy 2004).

Yet I show that this intuition can be misleading. In a work environment, advice helps an employee to solve a problem that jeopardizes her project outcome. The decision to ask for advice becomes more complex because additional effects on the employee's work must be taken into account. This novel approach results in three main findings that contradict the previous reasoning. First, the incentives to seek advice partly increase with the employee's ability. Secondly, advice-seeking can in fact signal high ability. Thirdly, reputation concerns can be beneficial; they increase advice-seeking in some situations.

The underlying reason for the model's findings is a positive, indirect effect of ability on advice-seeking that works through effort provision. First, a hard-working employee has higher incentives to seek advice because higher work effort raises expected benefits from a project and thus the need to solve the project-threatening problem. Secondly, a more competent employee works harder because ability increases the productivity of effort. Hence advice-seeking incentives increase with the seeker's ability. On the other hand, more competent employees are better problem-solvers themselves and hence have a lower need for advice which reflects the common intuition above. The two effects thus work in opposite directions. The overall relationship between ability and advice-seeking becomes non-monotonic. If the link between ability and effort is sufficiently pronounced, the overall incentives to seek advice increase with the seeker's ability.

Moreover, I show that reputation concerns may well have a positive effect on advice-seeking because they increase incentives to do so in some situations. There are two reasons to that. First, if project success is associated with high ability, reputation concerns make employees work harder. Consequently, as effort raises the incentives to seek advice, reputation concerns increase advice-seeking. Secondly, as outlined above, the incentives to seek advice can be higher for more competent employees. In that case, advice-seeking itself becomes a signal for high ability and employees start seeking advice to increase their reputation.

Chapter 1 suggests that the interaction between reputation and advice-seeking is more complex than previously thought, complementing a literature that has mainly focused on negative reputation effects of advice-seeking (Lee 2002; Levy 2004; Chandrasekhar et al. 2018). Yet the model includes these findings, precisely because the need for advice signals a lack of problem-solving skills and thus low ability. Further, due to the novel effort channel that arises in work environments reputation and advice-seeking are also positively linked. This new theoretical result is consistent with more recent evidence that highlights a potential positive interaction (Brooks et al. 2015; Thompson and Bolino 2018); there advice-seeking increases the seeker's reputation.

Chapter 2 investigates the optimal design of rules that are issued to promote prosocial behavior. That objective is especially crucial for organizations that are strongly committed to increase teamwork and cooperation. To this end they often rely on informal rules, for example encoded in employee handbooks or corporate statements. Rules are ubiquitous in organizations; however, they vary widely in scope, content and effectiveness. I explore how the design of such rules affects adherence to them. Thereby I compare the behavior and welfare implications of two common, but very distinct rules. The "Unconditional Rule" is universal in scope and content; it prescribes one action regardless of circumstances ("You shall not lie"). The "Conditional Rule" is very specific; it prescribes different actions for different situations ("If A, then do B").

I claim that the Conditional Rule can lead to less rule-following and lower welfare than the Unconditional Rule. That finding hinges on the existence of selfish individuals who want to be perceived as "good people", i.e. adherent to rules. It is known from previous literature that those individuals act more selfishly when their actions' observability decreases in a so-called "moral wiggle room" (Dana et al. 2007; Andreoni and Bernheim 2009). I first show theoretically that a similar mechanism results in less adherence to the Conditional Rule.

Since the Unconditional Rule prescribes only one action there is no uncertainty regarding an individual's intentions; hence if image concerns are sufficiently pronounced, even selfish individuals will follow the rule. The Conditional Rule, however, prescribes different actions for different situations. In an uncertain world, one can never be sure about an individual's intentions; either she acted selfishly on purpose, or she is a rule-follower who received a misleading signal. Therefore the observability of individual behavior decreases under the Conditional Rule. As a result, selfish, image-concerned individuals can break the rule and nonetheless maintain a positive image. A trade-off between the two rules emerges; the Unconditional Rule is inefficient in some situations, but followed by everyone, while the Conditional Rule is efficient in all situations, but only followed by individuals who always follow rules. Consequently the Unconditional Rule leads to higher welfare if too many people are indeed selfish in spirit.

Further I present results from a laboratory experiment to investigate rule-following behavior (i.e. contributions in a public goods game) under the two rules described above. The experiment shows that subjects tend to follow the rules. Furthermore, a moral wiggle room exists since selfish actions are deemed socially less inappropriate under the Conditional Rule. Nonetheless I do not find general differences in contribution behavior between the two rules. Yet the results also indicate that selfish subjects act more selfishly under the Conditional Rule if incentives to do so are sufficiently strong. In conclusion, the experimental evidence regarding moral wiggle rooms and behavior under rules remains inconclusive (in line with conflicting findings by Dana et al. 2007, Andreoni and Bernheim 2009 and van der Weele et al. 2014). Future research on the existence, reasons and implications of moral wiggle rooms is hence needed.

Chapter 3 directly deals with how to design an organization. More specifically, it investigates how a firm should optimally design its promotion rules and structure its decision-making process. It is concerned with the effects of promotions that are based on employee performance. In general, those promotions have two objectives, namely to motivate employees and to select good managers. However, empirical evidence by Benson et al. (2018) suggests that performance-based promotions in fact fail their second objective and select bad managers. This result resembles the "Peter Principle" (Peter and Hull 1969); promotions seem to promote the wrong employees. In this chapter I theoretically explain how such promotion policies can emerge from optimal organizational design.

The model is based on two premises. First, managers extract private benefits when they make decisions, at the firm's expense. Secondly, individuals differ in the amount of benefits they extract. Since performance-based promotions give rise to potential private benefits in the future, they motivate employees to work hard. However, due to the heterogeneity in benefit extraction employees are motivated differently by promotion prospects. Those employees who, as managers, extract more private benefits and decrease firm profits work harder and are thus more likely to get promoted.

Therefore performance-based promotions have two counteracting effects. They increase profits by motivating employees but the subsequent bad selection of managers has negative implications for firm profits. That trade-off affects the optimal design of organizations. I show that the joint use of partial delegation and performance-based promotions can be optimal. The principal uses promotions to motivate employees; however, he restricts a manager's decision rights to limit the private benefits extracted by the manager. Thereby he limits (a) the likelihood of promoting the "wrong" employee and (b) reduces the profit losses from the manager's rent extraction. Yet a negative selection effect still arises and the wrong employee is promoted with a higher probability. Chapter 3 complements a literature on the (negative) selection effects of promotions which focuses on skill-based explanations (Bernhardt 1995; Fairburn and Malcomson 2001); there performance-based promotions result in managers with wrong or worse skills. In contrast, I explain negative selection effects as an immediate consequence of performance-based promotions. They arise solely from the preference misalignment between manager and firm, and not the manager's skills.

In conclusion, this dissertation contains three contributions on the optimal design of organizations and its implications on individual behavior, in three distinct settings. It offers new insights into the mechanics behind the decision to seek advice, the optimality of rules to enhance prosocial behavior and the optimal design of promotion rules and delegation of authority.

Chapter 1

Advice-Seeking and Reputation

1.1 Introduction

> *"[H]elp seekers need help because of their inability to solve problems and find solutions on their own (...)"*
>
> *"By seeking help, one publicly acknowledges incompetence, inferiority, and dependence in front of another person, which can be highly threatening to one's public impressions within organizational settings."*

Lee (2002, p. 18/19)

Advice and help from other people are crucial for decision-making because an advisor can provide important information that the advice-seeker lacks. Indeed, advice and information lead to better decision-making, higher performance and higher profits.[1] But despite its positive effects individuals often refrain from actively asking for advice because they fear appearing incompetent (DePaulo and Fisher 1980). Accordingly, students do not ask for help from their teachers when they need it (Ryan et al. 2001). Employees do not seek advice either (Lee 1997), nor do they accept help that is offered to them (Thompson and Bolino 2018).

[1]For example, they improve financial decisions (Duflo and Saez 2003), educational choices (Jensen 2010), labor market outcomes (Altmann et al. 2018), creativity (Mueller and Kamdar 2011), and performance in experiments (Schotter 2003) as well as organizations (Podsakoff et al. 2009; Thompson and Bolino 2018).

A fear of reputation loss creates a tension between helpful advice-seeking and a high reputation. As Lee (2002) suggests, common intuition is that one only needs advice because of an inability to solve problems alone, which in turn implies low competence. Consequently, if a person is not seeking advice they signal not needing advice, and thus high competence. Given sufficiently strong reputation concerns, only those (incompetent) individuals with a very strong need for advice seek it. An inefficient separating equilibrium emerges as reputation concerns prevent people from seeking helpful advice (Levy 2004).

But this intuition is simplistic, for two reasons; first, reputation is built on more factors than only the decision to seek advice. For example, a teacher evaluates his students not only on whether they have asked for help, but on their exam performances. Similarly, an employee is assessed mainly on the basis of his or her (un)successful work. Secondly, this reasoning does not consider how the benefits of advice change with the advice-seeker's ability. For example competent students may study harder and thus benefit more from their teacher's help. In such environments advice-seekers must thus consider the *joint* effects of advice-seeking on their reputation, work performance and the interaction thereof. Consequently, the relationship between ability and advice-seeking becomes more nuanced than suggested.

This chapter explores the relationship between advice-seeking and reputation concerns in a work environment. I ask how advice-seeking and an employee's work effort interact and under what circumstances the employee seeks advice. Furthermore, the triangular relationship between ability, advice and effort is examined. I investigate whether the common intuition still holds true that only individuals with low ability need and thus seek advice. Lastly, I evaluate how reputation concerns and advice-seeking interact. Do reputation concerns necessarily prevent individuals from seeking helpful advice or are there positive reputation effects that eventually increase advice-seeking?

In search of answers, I model the decision problem of an agent who works on a project. At some point, a problem may appear which jeopardizes the project. To solve the problem the agent follows a step-by-step procedure. First, she tries to solve the problem on her own. If failing, she faces the decision of either (costly) seeking advice which solves the problem or trying to solve the problem on her own a second time. Ultimately, an unsolved problem leads to project failure; if the problem is solved, the project outcome depends on previous work effort.

The setup results in a complementarity between advice-seeking and effort provision. First, advice increases the productivity of effort and has a motivating effect on the agent. Without advice the problem may remain unsolved and all effort will be in vain. Hence, advice effectively "restores" previous work by solving the problem with certainty. Therefore, an agent who seeks advice works harder. Vice versa, higher effort increases the incentives to seek advice because it improves the chances of a successful project and thus the expected benefits from solving the problem increase. As advice affects the probability of problem-solving higher effort induces stronger incentives to seek advice.

Furthermore, I find that the relationship between ability and advice involves two counteracting effects that arise from a positive correlation between ability and problem-solving skills. First, a competent agent is less likely to need advice because of higher problem-solving skills. Therefore, higher ability means less advice-seeking. On the other hand, higher problem-solving skills also imply higher productivity of effort. Hence, competent agents work harder and as effort increases the incentives to seek advice, their likelihood to do so increases. Taken together, there are two counteracting forces. While the direct effect of ability on advice-seeking is negative, the indirect effect via effort provision implies a positive impact of ability on advice-seeking.

Who ultimately seeks advice depends on the relative strength of the two effects. If the complementarity between advice and effort is particularly pronounced, an "intermediate sorting" equilibrium arises. In that case, very competent agents refrain from seeking advice because they have a high probability of solving problems on their own. Also very *in*competent agents do not seek advice because their likelihood to need advice and thus bear additional advice-seeking costs is too high. For them, it is more profitable to work little and not seek advice. Therefore, only agents with intermediate ability seek advice.

Most empirical studies have focused solely on the negative relationship between ability and advice-seeking, investigating the seeker's fear of appearing incompetent (DePaulo and Fisher 1980; Ashford and Northcraft 1992; Lee 1997) and consequences in advice-seeking behavior (Lee 2002; Chandrasekhar et al. 2018). In contrast, some researchers investigate actual consequences on the seeker's reputation than just his beliefs, and find different results. In a series of lab experiments, Brooks et al. (2015) show that advisors have a higher perception of a subject's competence if they are asked for advice, in particular in difficult tasks. The result also applies when the advisors are given information about the subject's prior performance. In a survey study, Thompson and Bolino (2018) find that employees' negative beliefs about accepting help from co-workers is correlated with lower performance evaluations by their supervisors. Thus, accepting help implies higher reputation. Relatedly, Yoon et al. (2019) find that requesting time extensions for project work does not harm employees' reputation although the employees think so and thus refrain from requests for time extensions. These findings are consistent with a more nuanced relationship between ability and advice as it is outlined in this paper.

Therefore it is natural to re-examine the role of reputation concerns in advice-seeking decisions. Interestingly, I find a broader spectrum of equilibria than previous models. Indeed, my model incorporates negative effects of reputation as in Levy (2004) and Chandrasekhar et al. (2018), because agents fear signaling incompetence by seeking advice. However I also present equilibria in which reputation concerns induce more people to seek advice.

In this case, reputation concerns *increase* advice-seeking. The result emerges from two distinct mechanisms. First, reputation is built on more factors than just the decision to seek advice, for instance because it is also affected by the project outcome. If project success is associated with high ability, everyone works harder to signal competence. Due to the complementarity between effort and advice, the incentives to seek advice increase. Secondly, reputation concerns can increase advice-seeking because *advice-seeking signals high ability*. That finding is in direct contrast to the general fear of losing reputation when asking for advice. Its intuition relies on the aforementioned result that advice-seeking can increase with ability. If competent agents generally benefit more from advice-seeking but costs are too high, reputation concerns can be beneficial. If advice-seeking is a signal for high ability reputation concerns induce competent agents to seek advice. However they are not sufficiently pronounced to induce incompetent agents to seek advice; hence a separating equilibrium emerges, consistent with positive reputation effects of advice-seeking.

The model does not only account for the variety of empirical findings regarding advice-seeking, it is also applicable to a variety of different environments. As a leading example, I present the model in a work environment. The agent first works on a project when a problem may appear. If that is the case the agent can ask her supervisor for advice. However, the model also captures other circumstances such as a student who is preparing for her exam. Here comprehension problems can occur that threaten a successful exam, and then the student must ask her teacher for help. In that setting, the model finds that a teacher's accessibility will improve student performance. The teacher's help increases the probability of success by providing help *and* motivating his students to study harder. That intuition mirrors existing empirical evidence in education research (Zepke and Leach 2010; Johnson and LaBelle 2017)

Related Literature. Unlike in other social sciences, advice-seeking has received little attention in economics, with aforementioned exemptions of Levy (2004) and Chandrasekhar et al. (2018). Economists have been primarily interested in advice-taking and how advice influences decisions.[2] Therefore this work relates mostly to research on advice-seeking from psychology and management. I will provide a thorough overview over the advice-seeking literature in Section 1.5, after presenting the model and its results.

This work also relates to the literature on social image and reputation concerns. Bursztyn and Jensen (2017) provide an excellent overview over its empirical relevance. For theoretical work on that topic, I refer the interested reader to Bénabou and Tirole (2006) and Ellingsen and Johannesson (2008).

Advice-seeking is closely related to other concepts such as information-seeking, help-seeking and feedback-seeking. Thus, I like to clarify the interpretation of advice in the model. Advice enters the model in reduced form; it plainly solves a problem with certainty. To leave room for different interpretations of the advising process, the underlying problem-solving process is not modeled. For example, the advisor may literally advise (i.e. suggest a course of action that is then undertaken by the seeker), provide firsthand help (i.e. solve the problem himself) or give information to the seeker that

[2]Schotter (2003) provides an overview over four experimental studies on the effects of advice on individual behavior in the lab. Other work includes Croson and Marks (2001), Chaudhuri et al. (2008), and Brandts et al. (2015). For an overview of organizational psychology research on the effects of advice-taking, see Bonaccio and Dalal (2006).

helps to find a solution. Consequently, advice-seeking can also be interpreted as help-seeking or information-seeking and I sometimes use the terms help and advice interchangeably. Such interpretation is in line with the current literature (Brooks et al. 2015; van der Rijt et al. 2013). Yet it should be noted that the model does not capture the related subject of feedback-seeking. Feedback-seeking (Ashford and Cummings 1983) describes proactive search for information on previous performance and is not tied to a specific problem (van der Rijt et al. 2013). Even though reputation concerns also arise in such decisions, feedback-seeking is distinctive to the setting investigated here.

The chapter unfolds as follows. In Section 1.2 the model is introduced and the decision to seek advice is analyzed. Section 1.3 focuses exclusively on how the agent's ability affects her advice-seeking decision. Reputation concerns are introduced in Section 1.4. I first generally examine how reputation concerns affect the decision to seek advice before analyzing the relationship between reputation and advice-seeking in equilibrium. Section 1.5 contains an extended discussion. First, I discuss the main results in light of related literature as well as further implications that arise from the model. Secondly, I discuss open questions regarding advice-seeking in general. Section 1.6 concludes.

1.2 A Model of Advice-Seeking

I consider the decision problem of one risk-neutral agent (she) in a work environment. The agent faces two decisions, (1) how much effort $e \in [0,1]$ to provide, and (2) whether to seek advice or not. The latter decision is described by $a \in \{A, N\}$ where $a = A$ denotes advice-seeking.

The agent works on one project. A successful project has a value of v for the agent. If the project fails, the agent receives 0. At $t = 1$, she privately chooses an effort level $e \in [0,1]$. e determines the project's success probability via $prob(success|e) = e$ and comes at costs $\frac{c_e}{2}e^2$ with $c_e > 0$.

After the agent has chosen an effort level, a problem occurs with probability π at $t = 2$. If the problem is still unsolved when the project outcome is realized, the project fails with certainty. If the problem is solved by then, it is as if it never occurred and the project's success probability is still determined by previous effort provision e.

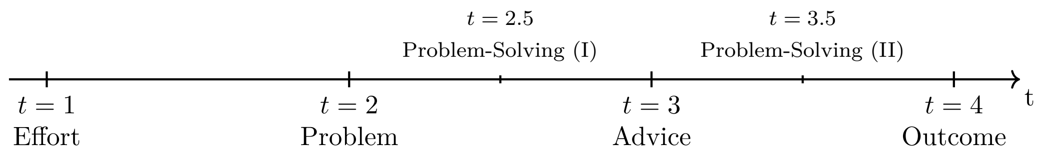


Figure 1.1: Advice-Seeking and Reputation - Timeline of the Model

In case a problem occurs, the agent first tries to solve the problem on her own at $t = 2.5$. Let $\lambda \in [0, 1]$ denote the probability that *she fails and the problem remains unsolved*.[3]

Given that she fails, advice is available with probability $\alpha \in [0, 1]$ at $t = 3$. If the agent seeks advice ($a = A$) she incurs advice-seeking costs of $c_A > 0$. Advice solves the problem with certainty. If the problem is unsolved at $t = 3.5$, because of not having sought advice or its unavailability, the agent tries to solve the problem a second time. Again, with probability λ she fails and the problem remains unsolved.

At $t = 4$ the game ends, the project outcome is realized and payoffs are made. If the problem is still unsolved, the agent receives a payoff of 0. If the problem has not occurred or was solved the project is successful with probability e, yielding a payoff of v and 0 otherwise. The model's full timeline is depicted in Figure 1.1. Since this constitutes a one-player game with complete information, I employ the subgame perfect Nash equilibrium as solution concept.

1.2.1 Advice-Seeking without Reputation Concerns

Let superscript A denote actions and utilities of an agent who seeks advice and N of an agent who does not. Call e^A the equilibrium effort provided by an advice-seeking agent, and e^N accordingly. Then, any subgame perfect Nash equilibrium in which the agent seeks advice is defined by the following two conditions. First, at the effort stage the agent provides effort e^A. Secondly, at the advice stage the agent prefers advice-seeking over no advice-seeking, given the previous effort choice e^A.

[3] Similar to its use in physics and engineering, λ thus describes the "failure rate" of an agent.

Formally the subgame perfect Nash equilibrium in which the agent seeks advice is hence characterized by

$$Eu^A_{t3}(e^A) \geq Eu^N_{t3}(e^A) \quad \text{and} \quad Eu^A_{t1}(e^A) \geq Eu^N_{t1}(e^N). \tag{1.1}$$

Likewise, the no-advice-seeking subgame perfect Nash equilibrium is characterized by $Eu^A_{t3}(e^N) \leq Eu^N_{t3}(e^N)$ and $Eu^A_{t1}(e^A) \leq Eu^N_{t1}(e^N)$.

The game is solved by backward induction. First, consider the advice stage, $t = 3$. Given effort e the agent receives an expected utility of $Eu^A_{t3} = e \cdot v - c_A$ if she seeks advice. If she does not, expected utility is given by $Eu^N_{t3} = (1 - \lambda) \cdot (e \cdot v) + \lambda \cdot 0$. Thus, at the advice stage the agent trades off expected benefits from advice-seeking $e \cdot \lambda v$ against certain costs c_A. She seeks advice if and only if

$$e \cdot \lambda v \geq c_A. \tag{1.2}$$

Advice-seeking changes the probability of success. If an agent does not seek advice, the project is successful with probability $(1 - \lambda) \cdot e$. Advice-seeking then increases the probability of success to $1 \cdot e$. That increase in expected utility, relative to the costs of advice-seeking, is captured in Equation (1.2).

At the effort stage, $t = 1$, the agent chooses an effort level e. In equilibrium, the agent anticipates her future advice-seeking choice. Thus, there are two optimal effort levels e^A and e^N, associated with anticipated advice-seeking and no advice-seeking in $t = 3$. Given that the agent seeks advice, she maximizes expected utility over e at $t = 1$ to find e^A:

$$\begin{aligned} \max_e \quad Eu^A_{t1} =& (1 - \pi\lambda)ev + \pi\lambda \cdot [\alpha(ev - c_A) + (1 - \alpha)(1 - \lambda)ev] - \frac{c_e}{2}e^2 \\ =& \left(1 - (1 - \alpha)\pi\lambda^2\right) ev - \frac{c_e}{2}e^2 - \pi\lambda\alpha \cdot c_A. \end{aligned} \tag{1.3}$$

The optimal effort when not seeking advice, e^N, is derived similarly:

$$\begin{aligned} \max_e \quad Eu^N_{t1} =& (1 - \pi\lambda)ev + \pi\lambda \cdot (1 - \lambda)ev - \frac{c_e}{2}e^2 \\ =& \left(1 - \pi\lambda^2\right) ev - \frac{c_e}{2}e^2. \end{aligned} \tag{1.4}$$

The optimal effort levels are then given by

$$e^A = \frac{(1-(1-\alpha)\pi\lambda^2)\,v}{c_e} \quad \text{and} \quad e^N = \frac{(1-\pi\lambda^2)\,v}{c_e}. \tag{1.5}$$

The comparative statics are straightforward. If a problem is not solved until the end of the game the project fails with certainty and all effort was in vain. Therefore, effort productivity decreases with the overall probability of not solving an occurring problem, given by the joint probability of a problem arising, π, and failing to solve the problem twice, λ^2. As a consequence, effort provision is decreasing in π and λ. Moreover, if the agent wants to seek advice the problem remains unsolved only if no advice is available. Hence, e^A is increasing in α and advice-seeking effort is more productive for any positive probability of receiving advice, i.e. $\alpha > 0$.

Lemma 1.1 shows the motivating effect of advice. Advice-seeking itself as well as the availability of advice increase effort provision. As advice increases chances of being successful, it motivates the agent to work hard if she seeks (and receives) advice if needed.

Lemma 1.1.
(1) An agent who seeks advice exerts higher effort than an agent who does not seek advice.
(2) Effort provision of an advice-seeking agent is increasing in the availability of advice.

Proof. The proof is relegated to Appendix A.1.

Lemma 1.1 is built on two observations that follow from (1.5). The first statement follows from $e^A > e^N \;\; \forall \alpha > 0$. The second statement follows from $\frac{\partial e^A}{\partial \alpha} > 0 \;\; \forall \alpha < 1$. Because e^N is independent of α, Lemma 1.1 also implies that the effort increase due to advice-seeking, i.e. $e^A - e^N$, is increasing in the availability of advice.

I proceed with the equilibrium analysis. Proposition 1.1 states the necessary and sufficient condition that determines advice-seeking in equilibrium.

Proposition 1.1.
In the unique subgame perfect Nash equilibrium, an agent seeks advice if and only if

$$\underbrace{\frac{\left(2-(2-\alpha)\pi\lambda^2\right)v}{2c_e}}_{\textit{Effort Effect}} \cdot \underbrace{\lambda v}_{\textit{Advice Effect}} \geq c_A. \qquad \text{(Condition I)}$$

The agent does not seek advice if and only if the inequality is reversed.

Proof. For a subgame perfect Nash equilibrium with advice-seeking, there are two conditions: $Eu^A_{t3}(e^A) \geq Eu^N_{t3}(e^A)$ and $Eu^A_{t1}(e^A) \geq Eu^N_{t1}(e^N)$. The first condition can be re-written as $e \cdot \lambda v \geq c_A$, see Equation (1.2). Expected utilities at $t = 1$ for the second condition are given by Equations (1.3) and (1.4). Plugging in the optimal effort levels given by Equation (1.5), I can re-write the two conditions as

$$Eu^A_{t3}(e^A) \geq Eu^N_{t3}(e^A) \quad \Leftrightarrow \quad \frac{\left(1-(1-\alpha)\pi\lambda^2\right)v}{c_e} \cdot \lambda v \geq c_A \tag{1.6}$$

$$Eu^A_{t1}(e^A) \geq Eu^N_{t1}(e^N) \quad \Leftrightarrow \quad \frac{\left(\left(1-(1-\alpha)\,\pi\lambda^2\right)v\right)^2}{2c_e} - \pi\lambda\alpha c_A \geq \frac{\left(\left(1-\pi\lambda^2\right)v\right)^2}{2c_e}. \tag{1.7}$$

First, simple mathematical reformulation yields $(2-(2-2\alpha)\pi\lambda^2)\cdot\lambda v^2 - 2c_e c_A \geq 0$ and $\alpha\pi\lambda \cdot \left((2-(2-\alpha)\pi\lambda^2)\lambda v^2 - 2c_e c_A\right) \geq 0$. That implies that Condition I is equivalent to the second condition and implies the first condition. Therefore it is necessary and sufficient. For the subgame perfect Nash equilibrium with no advice-seeking, the proof is similar and thus omitted here. □

To ease understanding, Condition I is decomposed into two effects. They show how expected benefits are formed by advice-seeking at the different stages of the game. The *Advice Effect* displays the gain from advice-seeking at $t = 3$. To see that, compare Condition I to Equation (1.2). An agent seeks advice if and only if $e \cdot \lambda v \geq c_A$. Here λv states the increase in expected benefits due to advice seeking. This is because the project can only be successful if the problem was solved before. While without advice-seeking the problem is solved with probability $(1-\lambda)$, advice solves the problem with certainty. The same intuition holds for the Advice Effect. It emphasizes the utility increase due to advice-seeking as advice solves the advice-seeker's problem.

The *Effort Effect* describes how the advice-seeking decision is shaped by effort considerations. It shows the complementarity between advice and effort that builds on Lemma 1.1. For illustration, the Effort Effect can be re-written as $\frac{(2-(2-\alpha)\pi\lambda^2)v}{2c_e} = e^N + \frac{e^A - e^N}{2}$. It becomes apparent that the Effort Effect is increasing in e^A and e^N and higher effort provision raises the incentives to seek advice. This is because advice-seeking "restores" effort by solving the problem. Hence, higher effort increases expected benefits from problem-solving, and thus advice-seeking. The Effort Effect further shows the motivating effect of advice-seeking, i.e. $\frac{e^A - e^N}{2}$, and its positive effect on expected utility. The associated increase in expected utility due to higher effort provision is adjusted by the costs of effort and divided in half.[4]

In Condition I, expected benefits are compared to costs of advice-seeking, given by c_A. Note that if $c_A = 0$, i.e. if advice is costless, the agent always seeks advice. However, there exist manifold reasons for positive advice-seeking costs, for instance opportunity costs or search costs to find suitable helpers (Hofmann et al. 2009).

1.2.2 Comparative Statics

The Effect of Advice Availability α. An increase in advice availability increases advice-seeking since the left-hand side of Condition I increases with α. This is due to the Effort Effect. Lemma 1.1 shows that the difference in effort provision, $e^A - e^N$, is increasing in advice availability. If advice becomes more likely effort productivity increases. Hence the availability of advice increases effort provision and ultimately the incentives to seek advice.

The Effect of Problem Probability π. The effect of a higher problem probability on advice-seeking is negative since the left-hand side of Condition I decreases with π. Again, the intuition comes from the Effort Effect. A higher problem probability demotivates the agent (and thus decreases both e^A and e^N). Therefore, at the advice stage the agent has little at stake because the probability of project success is low. Consequently incentives to seek advice decrease with π.

The comparative statics with respect to λ, the probability of solving problems, are investigated in more detail in the next section. Further assuming a positive correlation between λ and the agent's ability allows for a better interpretation of the problem-solving process.

[4]This is an immediate consequence of the model's functional forms, i.e. the linear success probability and convex costs.

1.3 Who Seeks Advice? The Role of Ability

I impose that higher ability correlates with better problem solving skills.[5] Denote θ as the agent's ability. Let $\lambda = \lambda(\theta) \in [0, 1]$ be decreasing in θ, i.e. $\frac{\partial \lambda(\theta)}{\partial \theta} < 0$. Hence if ability θ increases, the probability of solving a problem, i.e. $1 - \lambda$, is increasing as well. Due to the one-to-one relationship between θ and λ and for notational ease, I will drop θ and refer to the agent's type merely as λ. It is important to note that a high λ corresponds to low ability, and vice versa.

In what follows I first analyze how expected benefits from advice-seeking are affected by the agent's ability in Lemma 1.2. In fact, I show that that the relationship between advice-seeking and ability is non-monotonic.

Lemma 1.2.
Define $f(\lambda) := (2 - (2 - \alpha)\pi\lambda^2)\lambda v^2 - 2c_e c_A$. $f(\lambda)$ *is concave in* λ, *with* $\overline{\lambda} = \sqrt{\frac{2}{3(2-\alpha)\pi}}$ *as its unique global maximum.*

(a) *If* $\lambda < \overline{\lambda}$, *the expected gains from advice-seeking decrease with ability, i.e.* $\frac{\partial f(\lambda)}{\partial \lambda} > 0$.

(b) *If* $\lambda > \overline{\lambda}$, *the expected gains from advice-seeking increase with ability, i.e.* $\frac{\partial f(\lambda)}{\partial \lambda} < 0$.

Proof. The proof is relegated to Appendix A.1.

To understand why expected benefits from advice-seeking are concave in λ, first note that $f(\lambda)$ is a re-formulation of Condition I. If and only if $f(\lambda) \geq 0$, the expected net benefits from advice-seeking are positive. Hence, I can analyze the relationship between ability and advice by examining the functional form of $f(\lambda)$. As f is concave in λ so are expected benefits from advice-seeking. For the intuition, I examine the impact of λ on the Advice Effect and the Effort Effect separately.

A higher λ, i.e. lower ability, increases the Advice Effect. If λ increases, the probability to solve a problem after not seeking advice decreases. Hence,

[5] That assumption is consistent with the definition of ability or competence in psychology. Gardner (1983, p.13) argues that "a human intellectual competence must entail a set of skills of problem solving — enabling the individual to resolve genuine problems or difficulties that he or she encounters and, when appropriate, to create an effective product (...)".

the relative benefits of seeking advice increase in λ. Individuals with low ability who are bad at solving problems gain more from seeking advice. It follows that the Advice Effect captures a *negative* effect of ability on the incentives to seek advice.

On the other hand, a higher λ decreases effort provision. This is because agents with high ability are more likely to solve the problem on their own. Therefore, their expected productivity of effort is higher and thus high ability types work harder. On the opposite, agents with low ability have a higher likelihood of project failure because of an unsolved problem. In that case, all effort was in vain. Agents with a low ability are demotivated by this prospect and exert lower effort. Therefore ability has a positive effect on effort provision. Condition I shows a complementarity between effort and advice. Consequently, since ability increases effort provision and higher effort makes advice-seeking more profitable the Effort Effect captures an indirect *positive* effect of ability on the incentives to seek advice.

Thus, the impact of λ on advice-seeking consists of two counteracting forces. The Advice Effect induces a negative effect of ability on advice-seeking via problem solving. It captures the common intuition that individuals with a lower ability are more likely to seek advice because they cannot solve problems on their own due to their low ability. Thus, they have a higher demand for advice. Secondly, the Effort Effect provides a foundation for a positive effect of ability on advice-seeking via effort provision.

Both effects taken together result in a concave relationship between ability and advice. The Effort Effect gives a natural "upper bound" to the negative impact of ability on advice-seeking incentives that arises from the Advice Effect. How behavior is ultimately affected by the agent's ability then hinges upon the relative strengths of the two effects. Before proceeding with Proposition 1.2 that summarizes all potential advice-seeking equilibria, I state in Lemma 1.3 that for any advice-seeking to take place, its costs must be sufficiently low.

Lemma 1.3.
If $c_A > \overline{c}_A = \frac{2v^2}{3c_e}$, *the agent never seeks advice regardless of her type.*

Proof. The proof is relegated to Appendix A.1.

The intuition is straightforward. If costs of advice-seeking are too high the agent never has an incentive to seek advice at $t = 3$. Lemma 1.3 gives an upper bound $\overline{c}_A$ to those advice-seeking costs. If they exceed $\overline{c}_A$, even the maximum expected benefits from advice-seeking are not sufficient to induce any type to seek advice. With that observation, I turn to the effects of ability on advice-seeking in equilibrium. To this end it is useful to define λ_1 and λ_2 as the positive roots of $f(\lambda) = (2 - (2 - \alpha)\pi\lambda^2)\lambda v^2 - 2c_e c_A$, if they exist.

Proposition 1.2.
Suppose $c_A \leq \overline{c}_A$. There exists a unique subgame perfect Nash equilibrium in which the agent exerts effort according to (1.5) *and her advice-seeking behavior is as follows.*

(a) "Negative sorting": If $\pi \leq \frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2}$, an agent of type λ seeks advice if and only if

$$\lambda_1 \leq \lambda \leq 1. \tag{1.8}$$

(b) "Intermediate sorting": If $\frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2} < \pi \leq \frac{2}{3(2-\alpha)}(\frac{2v^2}{3c_e c_A})^2$, an agent of type λ seeks advice if and only if

$$\lambda_1 \leq \lambda \leq \lambda_2 < 1. \tag{1.9}$$

(c) "No Advice-Seeking": If $\frac{2}{3(2-\alpha)}(\frac{2v^2}{3c_e c_A})^2 < \pi$, the agent never seeks advice regardless of her type.

Proof. Recall $f(\lambda) := ((2 - (2 - \alpha)\pi\lambda^2)\lambda v^2) - 2c_e c_A$ with $\lambda \in [0, 1]$. Thus, if and only if $f(\lambda) \geq 0$, Condition I holds. Therefore I am interested in the roots of $f(\lambda)$.

First, note that at $\lambda = 0$, $f(\lambda) = -2c_e c_A < 0$ and $f'(\lambda) = 2v^2 > 0$. Therefore, no advice-seeking takes place at $\lambda = 0$. Secondly, note that if $f(\overline{\lambda}) < 0$, there is no advice-seeking because even the type with the highest benefits from advice-seeking, $\overline{\lambda}$, does not seek advice. Using $\overline{\lambda} = \sqrt{\frac{2}{3(2-\alpha)\pi}}$, $f(\overline{\lambda}) < 0$ gives an upper bound $\overline{\pi} = \frac{2}{3(2-\alpha)}(\frac{2v^2}{3c_e c_A})^2$ on the problem probability. Only if the probability of a problem arising is sufficiently low, i.e. if $\pi \leq \overline{\pi}$, advice-seeking may take place.

Keeping that in mind, turn to the analysis of $f(\lambda)$'s roots. Due to the concavity of f and $f(0) < 0$, there are between zero and two roots for f. Remember that λ_1 and λ_2 denote f's potential roots, with $\lambda_1 < \lambda_2$.

First, I investigate the case of zero roots. Because f is concave in λ and $f(\lambda = 0) < 0$ it is apparent that f has zero roots if at $\lambda = 1$ it is negative but still increasing. Therefore, f has zero roots if $f(\lambda = 1) < 0$ and $f'(\lambda = 1) > 0$. We can rewrite these conditions as $\pi > \frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2}$ and $\pi < \frac{2}{3(2-\alpha)}$. In that case, f has never crossed the abscissa, therefore $f(\lambda) < 0 \; \forall \lambda$ and no agent ever seeks advice.

f has one root if and only if $f(\lambda = 1) \geq 0$. Then, f has crossed the abscissa only once, at λ_1. We can rewrite this condition as $\pi \leq \frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2}$. In this case, it holds that $f(\lambda) < 0$ for $\lambda < \lambda_1$ and $f(\lambda) \geq 0$ for $\lambda \geq \lambda_1$. This corresponds to a "negative sorting" equilibrium. Only types with sufficiently low ability that corresponds with a sufficiently high λ seek advice.

There are two roots if, at $\lambda = 1$ f is negative but decreasing and $f(\overline{\lambda}) > 0$. The last condition will be examined at the end of the proof, for now suppose it holds true. The first two conditions are captured by $f(\lambda = 1) \leq 0$ and $f'(\lambda = 1) < 0$. We can rewrite these conditions as $\pi \geq \frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2}$ and $\pi > \frac{2}{3(2-\alpha)}$. In this case, $f(\lambda) > 0$ if and only if $\lambda_1 \leq \lambda \leq \lambda_2 \leq 1$. This corresponds to the "intermediate sorting" equilibrium. Only types of an intermediate λ that corresponds with $\lambda_1 \leq \lambda \leq \lambda_2$ seek advice.

Together with $\pi \leq \overline{\pi} = \frac{2}{3(2-\alpha)}(\frac{2v^2}{3c_e c_A})^2$, I can summarize as follows.

(1) There is no advice-seeking if $\pi > \overline{\pi}$ or $\frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2} < \pi < \frac{2}{3(2-\alpha)}$.

(2) There is advice-seeking for all $\lambda \geq \lambda_1$ if $\pi \leq \min\left\{\frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2}, \overline{\pi}\right\}$.

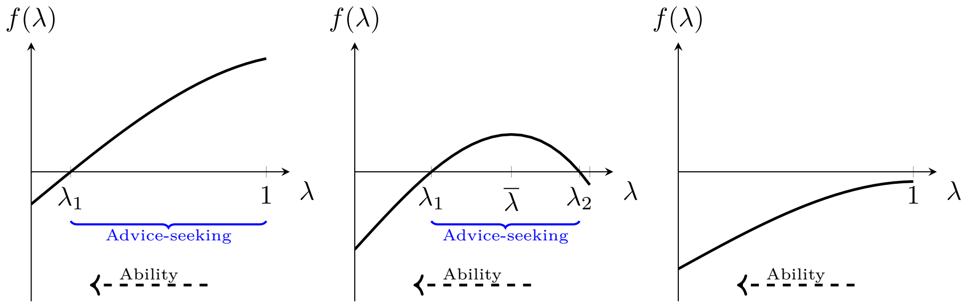
(3) There is advice-seeking for all $\lambda \in [\lambda_1, \lambda_2]$ if $\max\left\{\frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2}, \frac{2}{3(2-\alpha)}\right\} < \pi < \overline{\pi}$.

As a next step, $c < \overline{c}_A$ implies $\frac{2v^2}{3c_e c_A} > 1$. Therefore, $2/3(2-\alpha) < (2v^2/3c_e c_A)^2 \cdot (2/3(2-\alpha)) = \overline{\pi}$. Hence, the second condition of case (1) is unfeasible and the condition on case (3) can be simplified to $\frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2} < \pi \leq \frac{2}{3(2-\alpha)}(\frac{2v^2}{3c_e c_A})^2$, as stated in Proposition 1.2.

Secondly, $f(\lambda = 1) > 0$ implies $f(\overline{\lambda}) > 0$. Therefore, $\pi \leq \frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2}$ implies $\pi < \frac{2}{3(2-\alpha)}(\frac{2v^2}{3c_e c_A})^2 = \overline{\pi}$. It follows that $\frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2} < \overline{\pi}$. That simplifies the condition in the second case to $\pi \leq \frac{2(v^2 - c_e c_A)}{(2-\alpha)v^2}$, as stated in Proposition 1.2. □

The expected net benefits of advice-seeking are sketched in Figure 1.2 and illustrated by $f(\lambda)$. Condition I holds if and only if $f(\lambda) \geq 0$ because $f(\lambda)$ is a re-formulation of it. To examine the influence between ability and advice-seeking, I focus on how $f(\lambda)$ reacts to changes in λ as displayed in Figure 1.2. Furthermore, recall the negative relationship between ability and the probability of not solving the problem, λ. The closer λ is to zero, the higher the agent's ability.

First, Figure 1.2 shows that agents with a very high ability (low λ) do not seek advice as $f(\lambda)$ is negative. For them, it is never worthwhile to seek advice because their probability of solving the problem alone is sufficiently high. Therefore, they never have an incentive to bear the costs of advice-seeking c_A. Secondly, all graphs are concave in λ. As discussed before, that



(a) Negative Sorting (b) Intermediate Sorting (c) No Advice-Seeking

Figure 1.2: The Influence of Ability on Advice-Seeking

Note: This is an illustration of Proposition 1.2. Note that ability increases as λ decreases as displayed by the dashed arrow.

concavity hinges on the relative size of the Advice Effect and the Effort Effect; the stronger the Effort Effect, the more the advice-seeking incentives react to changes in ability. Proposition 1.2 gives conditions on the relative size of the Effort Effect in terms of the problem probability, π. Thus π determines what equilibrium emerges. To see why, note that it has two effects on the decision to seek advice. First, it decreases the Effort Effect as a higher problem probability decreases incentives to exert high effort (see Lemma 1.1). Secondly, it increases the marginal effect of ability on the Effort Effect.

Figure 1.2 presents the three potential equilibria described in Proposition 1.1. In Figure 1.2a, π is relatively small and only types with a sufficiently low ability seek advice. Note that a small problem probability π implies that the absolute size of the Effort Effect is high and so are the incentives to seek advice. Therefore, the advice-seeking threshold λ_1 is relatively low. Furthermore, a small π also implies that ability only weakly affects effort, displayed as the low marginal effect of λ on the slope of $f(\lambda)$. Therefore, the positive effect of ability on advice via effort provision is not particularly pronounced. As a result, the (negative) impact of ability on advice-seeking via the Advice Effect becomes superior. Hence, the incentives to seek advice monotonically increase with λ and the negative sorting equilibrium arises. Only low ability types (with $\lambda \geq \lambda_1$) seek advice.

In Figure 1.2b, π is intermediate. Compared to Figure 1.2a, the absolute size of the Effort Effect has decreased as effort is reduced when π increases. Therefore, the agent's demand for advice must increase to make her seek advice. Consequently, the threshold λ_1 increases compared to Figure 1.2a. A higher π also results in a stronger positive effect of ability on advice-seeking incentives. Therefore, the Effort Effect decreases strongly if λ increases and the positive effect is superior for all $\lambda > \overline{\lambda}$. In that case, the incentive to seek advice *decreases* with λ; thus a very incompetent agent ($\lambda > \lambda_2$) has no incentive to seek advice. The intermediate sorting equilibrium emerges.

In Figure 1.2c, π is very high. As a consequence, the agent is strongly demotivated to exert effort because of the likely prospect of project failure, irrespective of her ability. Therefore, the absolute value of the Effort Effect is very low and at the advice stage, no type is willing to bear the costs of seeking advice. Hence the no advice-seeking equilibrium arises.

To sum up, the effect of ability on advice-seeking is two-fold. On the one hand, the most competent individuals have no incentive to seek advice because they are sufficiently good at problem-solving themselves. Yet, as ability decreases the incentives to do so increases. On the other hand, as outlined in this section, effort decreases with ability which leads to a counteracting effect. Three equilibria may emerge: negative sorting, intermediate sorting and no advice-seeking. That finding also changes the role of reputation concerns; if mainly high types seek advice, reputation concerns may induce others to seek advice as well.

1.4 Who Seeks Advice? The Role of Reputation

Reputation concerns affect voting behavior, educational choices, effort in the workplace or financial investments (for an overview, see Bursztyn and Jensen 2017). Ample evidence further emphasizes its importance in the advice-seeking decision. Indeed, people fear a reputation loss because of signaling incompetence when asking for advice (Lee 2002). This idea is in line with the previously found negative sorting equilibrium. Since the Advice Effect is particularly pronounced, only agents with low ability seek advice. A neutral observer thus infers low ability if someone seeks advice. However, the Effort Effect gives rise to an equilibrium in which the most incompetent also do not seek advice. In that case, the intermediate sorting equilibrium emerges.

Previous models on the interplay between advice-seeking and reputation concerns have focused solely on the former intuition (Levy 2004; Chandrasekhar et al. 2018). They find that reputation concerns are harmful and induce inefficiencies because they prevent individuals from advice-seeking. But with the findings from Section 1.3 in mind, it is natural to re-examine this relationship. In this section I show that the model encompasses two conflicting views on the relationship between advice-seeking and reputation.

The negative view on reputation effects, described in Section 1.4.3, is consistent with the fear of reputation loss and subsequently fewer people seeking advice. Since only incompetent agents seek advice, it reduces reputation. As a consequence, fewer people ask for advice because they fear signaling incompetence. Therefore, the model captures previous findings on advice-seeking that have focused on that fear (for a thorough discussion, see Section 1.5.2).

The positive view, described in Section 1.4.3, highlights a different effect of reputation concerns. In particular, it builds on the positive effect of ability on effort provision. There are two different mechanisms how that influences the role of reputation concerns. First, high types exert higher effort; hence successful projects become a signal for high ability (Theorem 1.1). Consequently reputation concerns increase all types' effort provision. Because effort and advice are complementary, reputation concerns ultimately increase advice-seeking even though it remains a signal for low ability.

Furthermore, Theorem 1.2 presents an equilibrium in which advice-seeking in fact even signals high ability. This is not only contrary to previous findings but also to the common intuition that "by seeking help one publicly acknowledge incompetence" (Lee 2002, p. 19). In the previous section I show that agents with higher ability have higher incentives to seek advice (in the "intermediate sorting" equilibrium). Theorem 1.2 builds on that intuition. There, I construct a separating equilibrium in which only high types have a strong incentive to seek advice and are induced by reputation concerns to do so, hence advice-seeking signals high ability.

This section builds on the previous model, with few adjustments made and presented in Section 1.4.1. I first investigate the general incentives to seek advice under reputation concerns in Section 1.4.2. The equilibrium effects of reputation concerns on advice-seeking are then examined in Sections 1.4.3.

1.4.1 Adjustments to the Model

To focus on the effects of reputation concerns I make the following adjustments. First, advice is always available, i.e. $\alpha = 1$. Secondly, the agent's ability and effort are private; thus the game becomes one of incomplete information. Thirdly, the agent is one of two ability types, a low type θ_l or a high type $\theta_h > \theta_l$. The prior is given by $prob(\theta = \theta_h) = \mu$. Since the high type is of higher ability she is more likely to solve a problem. Denote the types' respective probability of failing to solve the problem by λ_l and λ_h, respectively. Hence $\lambda_l > \lambda_h$. Note that λ_l denotes the low (ability) type even though $\lambda_l > \lambda_h$. As before, I will use $\lambda \in \{\lambda_l, \lambda_h\}$ to refer to an agent's type for notational ease.

The agent's type λ and the chosen effort level e are private information. Her decision to seek advice $a \in \{A, N\}$ and the project outcome $y \in \{F(ail), S(uccess)\}$ are publicly observable.

The agent cares about her reputation, i.e. an observer's posterior belief $\hat{\mu}(a, y)$ after observing (a, y) and using Bayes' rule. The exact functional form of the agent's reputation utility R is assumed to be

$$R(\hat{\mu}(a, y), \overline{\mu}, r) = \begin{cases} r & \text{if } \hat{\mu}(a, y) \geq \overline{\mu} \\ 0 & \text{if } \hat{\mu}(a, y) < \overline{\mu}. \end{cases} \tag{1.10}$$

This reputation utility function is a step function. If the observer's posterior belief exceeds a threshold $\overline{\mu}$ the agent receives a reputation utility of r.

This functional form implies that gains from reputation emerge from a binary choice. Such a setting can be found in many instances, for example promotions in organizations. Suppose an employee's supervisor is more knowledgeable and can solve problems better than an agent. When an employee cannot solve her problem alone she turns to her supervisor for advice.[6]

Yet an employee also cares about her reputation because the supervisor decides over relevant outcomes such as bonuses or promotions. As Prender-

[6] This process is common in organizations. Garicano (2000) gives a theoretical foundation of why knowledge is concentrated in higher tiers. Also in his model, if an employee cannot solve a problem alone, it is passed on to his direct supervisor. On the empirical side, Fisher et al. (2018) find that the majority of employees go to their direct bosses to ask for help. The interpretation of the supervisor as advisor is also consistent with a small literature on expert leadership (Goodall et al. 2011).

gast (1999, p. 33) states, "most workers in the economy are evaluated subjectively". Moreover, Frederiksen et al. (2017) and Frederiksen et al. (2019) show how a supervisor's subjective rating of his supervisee's performance has strong effects on the supervisee's career outcomes. In that setting, one can interpret r as expected gains from a future promotion. If the supervisor bases the binary promotion decision on his subjective beliefs about the agent's ability, reputation concerns arise. Before seeking advice from a supervisor one must consider the reputation effects of such action.

1.4.2 Advice-Seeking with Reputation Concerns

I first re-examine the agent's decision to seek advice in the light of reputation concerns before investigating the equilibrium effects. For notational ease it is useful to define $R^{a,y} := R(\hat{\mu}(a, y), \overline{\mu}, r)$ as the agent's reputation utility if the observer's posterior belief is based on advice-seeking decision a and project outcome y. Furthermore, define $d^a := R^{a,S} - R^{a,F}$ as the advice-seeking specific differences in reputation utility dependent on the project outcome. Note that both $R^{a,y}$ and d^a depend on the action $a \in \{A, N\}$ as well as the associated posterior beliefs that emerge in equilibrium and are therefore dependent on the equilibrium under consideration. Yet I first analyze general implications of reputation concerns (thus treat R and d as exogenous) before considering the equilibrium effects of reputation in Section 1.4.3.

The equilibrium concept is Perfect Bayesian Equilibrium (PBE). In equilibrium, advice-seeking needs to be optimal at the effort and advice stage. As in the previous analysis without reputation concerns, in any PBE an agent seeks advice if and only if

$$E\tilde{u}^A_{t1}(e^A) \geq E\tilde{u}^N_{t1}(e^N) \quad \text{and} \quad E\tilde{u}^A_{t3}(e^A) \geq E\tilde{u}^N_{t3}(e^A). \tag{1.11}$$

To find the necessary and sufficient condition for advice-seeking in the presence of reputation concerns I proceed as in Section 1.2. First, I write expected utilities at the advice stage given any previous effort provision e as

$$E\tilde{u}^A_{t3} = e(v + d^A) + R^{A,F} - c_A \tag{1.12}$$

$$E\tilde{u}^N_{t3} = (1 - \lambda) \cdot e(v + d^N) + R^{N,F}. \tag{1.13}$$

Thus, the agent seeks advice at $t = 3$ if and only if

$$e \cdot \lambda v + R^{A,F} - R^{N,F} + e \cdot (d^A - (1-\lambda)d^N) \geq c_A. \tag{1.14}$$

This expression is similar to the one without reputation concerns, see Equation (1.2). Comparing the two, I find that the project-related benefits of advice-seeking are $e \cdot \lambda v$ in both cases. The reputation effects of advice-seeking depend on the project outcome. If the project fails the difference in the agent's utility is given by $R^{A,F} - R^{N,F}$. If the project succeeds there are additional changes in the agent's reputation gains. These are conditional on advice-seeking (i.e. d^A and d^N). As before, the project succeeds with probability e if the agent seeks advice, but only with probability $e(1-\lambda)$ if the agent does not seek advice. Thus, the agent's expected reputation gains from a successful project are $e \cdot (d^A - (1-\lambda)d^N)$.

In the next step, I determine the optimal effort levels $\tilde{e}^A$ and $\tilde{e}^N$. The respective expected utilities at $t = 1$ are maximized w.r.t. effort level e:

$$\max_e \; E\tilde{u}^A_{t1} = e \cdot (v + d^N) + \pi\lambda \cdot [x(d^A - d^N) + R^{A,F} - R^{N,F}] - \pi\lambda c_A - \frac{c_e}{2}e^2 \tag{1.15}$$

$$\max_e \; E\tilde{u}^N_{t1} = (1 - \pi\lambda^2) \cdot e \cdot (v + d^N) - \frac{c_e}{2}e^2. \tag{1.16}$$

The optimal effort levels with reputation concerns are given by

$$\tilde{e}^A = \frac{v + d^N + \pi\lambda(d^A - d^N)}{c_e} \quad \text{and} \quad \tilde{e}^N = \frac{(1-\pi\lambda^2)(v + d^N)}{c_e}. \tag{1.17}$$

Lemma 1.4 presents the comparative statics with regard to reputation concerns in equilibrium. First, if project success increases reputation (i.e. $d^A > 0, d^N > 0$) the agent exerts higher effort, regardless of the advice-seeking decision. The intuition is straightforward; reputation concerns increase the benefits of a successful project and thus motivative the agent. But if the reputation utility is especially pronounced when not seeking advice (i.e. d^N is sufficiently large), an advice-seeking agent may work less. Then the reputation incentives to work hard are low for advice-seekers while an agent who does not seek advice is particularly motivated by their reputation concerns.

Lemma 1.4.
(1) Effort increases with reputation concerns if in equilibrium $d^A > 0$ *and* $d^N > 0$*, regardless of the advice-seeking decision.*
(2) An agent who seeks advice exerts higher effort than one who does not seek advice, i.e. $\tilde{e}^A > \tilde{e}^N$*, if in equilibrium* $d^N < \frac{\lambda v + d^A}{1-\lambda}$.

Proof. The proof is relegated to Appendix A.1.

I proceed with the necessary and sufficient condition for advice-seeking in equilibrium with reputation concerns in Proposition 1.3. It expands Proposition 1.1 by the additional effects of reputation on advice-seeking.

Proposition 1.3.
In any Perfect Bayesian Equilibrium with reputation concerns, an agent of type λ *seeks advice if and only if*

$$\underbrace{\frac{(2-\pi\lambda^2)(v+d^N)+\pi\lambda(d^A-d^N)}{2c_e}}_{\textit{Effort Effect}} \cdot \underbrace{[\lambda v + d^A - (1-\lambda)d^N]}_{\textit{Advice Effect}} + \underbrace{(R^{A,F} - R^{N,F})}_{\textit{Pure Reputation}} \geq c_A. \tag{Condition II}$$

Proof. The proof is similar to the one of Proposition 1.1 and is thus omitted here. In general, one can show that Condition II is equivalent to $E\tilde{u}^A_{t1}(e^A) \geq E\tilde{u}^N_{t1}(e^N)$ and implies $E\tilde{u}^A_{t3}(e^A) \geq E\tilde{u}^N_{t3}(e^A)$.

Condition II is decomposed into three different effects. Besides the Effort Effect and the Advice Effect, already known from the analysis without reputation concerns, a third effect emerges, the "Pure Reputation Effect". The interpretation of the Effort Effect and the Advice Effect remain unchanged. Advice-seeking increases the probability of project success and thus expected utility, as shown by the Advice Effect. The Effort Effect describes the complementarity between effort and advice. Lastly, the Pure Reputation Effect shows the reputation benefits of advice-seeking without any effort provision (i.e. in case of certain failure of the project).

Reputation concerns influence the decision to seek advice through all three channels. First, the Pure Reputation Effect arises solely from reputation concerns. It captures the advice-seeking difference in reputation conditional on project failure. Secondly, the introduction of reputation concerns changes the Advice Effect, i.e. the expected relative payoff if the project is successful, by $d^A-(1-\lambda)d^N$ (see the discussion on Equation (1.14)). Thirdly, reputation concerns affect effort provision, and thus the Effort Effect, via $\tilde{e}^A$ and $\tilde{e}^N$. Lemma 1.4 shows that effort provision increases if project success is associated with a high type. That intuition can be applied here as well: if project success is associated with a high type, i.e. $d^A > 0, d^N > 0$, the Effort Effect and thus the incentives to seek advice increase.

1.4.3 The Effects of Reputation Concerns

As in any signaling game, posterior beliefs are determined only in equilibrium; thus the exact implications of reputation concerns depend on equilibrium inferences. Therefore I analyze the equilibrium effects of reputation concerns on advice-seeking in this section. I focus solely on the comparison of advice-seeking behavior with and without reputation concerns.[7] Hereby, I investigate effects on the extensive margin and compare advice-seeking behavior of an agent with type $\lambda \in \{\lambda_l, \lambda_h\}$ without reputation concerns and with reputation concerns, given $(r, \overline{\mu})$. More specifically, define

$$\begin{aligned}\tilde{f}(\lambda) :=& [(2-\pi\lambda^2)(v+d^N)+\pi\lambda(d^A-d^N)]\cdot[\lambda v + d^A-(1-\lambda)d^N] \\ &+ 2c_e(R^{A,F}-R^{N,F}) - 2c_e c_A. \qquad (1.18)\end{aligned}$$

Here, posterior beliefs are determined in equilibrium and reputation utility realizes according to the posterior beliefs and $(r, \overline{\mu})$. As $\alpha = 1$ is imposed throughout this section, redefine $f(\lambda)$ as

$$f(\lambda) := (2-\pi\lambda^2)\lambda v^2 - 2c_e c_A. \qquad (1.19)$$

As before, an agent of type $\lambda \in \{\lambda_l, \lambda_h\}$ seeks advice without reputation concerns if and only if $f(\lambda) \geq 0$. When her reputation is involved, she seeks advice if and only if $\tilde{f}(\lambda) \geq 0$. Throughout the analysis I focus on the *extensive margin* of reputation effects. Thereby, I compare actual advice-seeking behavior with and without reputation concerns and do not consider

[7]For the interested reader, I conduct a full equilibrium analysis of all Perfect Bayesian equilibria with reputation concerns in pure strategies in Appendix A.2.

marginal changes, neither of reputation nor advice-seeking incentives. I define positive and negative effects of reputation concerns as follows.

1. Reputation concerns have a positive effect on advice-seeking if there exists at least one type λ who is induced to switch from no advice-seeking to advice-seeking but not vice versa. Formally,

$$\exists\, \lambda \in \{\lambda_l, \lambda_h\} : f(\lambda) < 0 \;\wedge \tilde{f}(\lambda) \geq 0$$
$$\nexists\, \lambda \in \{\lambda_l, \lambda_h\} : f(\lambda) \geq 0 \;\wedge \tilde{f}(\lambda) < 0.$$

2. Reputation concerns have a negative effect advice-seeking if there exists at least one type λ who seeks advice without reputation concerns but does not with reputation concerns. Formally,

$$\exists\, \lambda \in \{\lambda_l, \lambda_h\} : f(\lambda) \geq 0 \;\wedge \tilde{f}(\lambda) < 0$$
$$\nexists\, \lambda \in \{\lambda_l, \lambda_h\} : f(\lambda) < 0 \;\wedge \tilde{f}(\lambda) \geq 0.$$

On the Negative Effects of Reputation Concerns

In the following analysis I will show the existence of the two different effects of reputation concerns. I start with the negative effects and examine the existence of a PBE in the spirit of Chandrasekhar et al. (2018). In their setting, everyone seeks advice without reputation concerns. Reputation concerns then lead to a negative sorting equilibrium. Above some cutoff, high ability types stop seeking advice because they fear for their reputation. Such an equilibrium is also embedded in the current model. It is described in Proposition 1.4 and follows a similar intuition. Moreover, I show in Proposition 1.5 that reputation concerns can even lead to a full termination of advice-seeking. In that case, no type seeks advice due to a fear of signaling a low type.

Proposition 1.4 presents an equilibrium in which both types pool on advice-seeking without reputation concerns but separate with reputation concerns and only the low type seeks advice. In the latter case advice-seeking becomes fully informative and reveals a low type. Hence there is no reputation utility when seeking advice, i.e. $R^{A,S} = R^{A,F} = 0$; but an agent who does not seek advice signals a high type. In the equilibrium described by Proposition 1.4, no advice-seeking thus leads to reputation gains, i.e. $R^{N,S} = R^{N,F} = r$.

Because advice-seeking is fully informative the project outcome bears no implications for the seeker's reputation. Hence effort is not affected by reputation concerns. Since the low type seeks advice in equilibrium, she exerts higher effort which in turn increases her incentives to seek advice further.

Therefore, the two types face different trade-offs at the advice stage. The low type has more at stake and a lower probability of solving the problem. Hence she accepts a low reputation and rather seeks advice to solve the problem with certainty. The high type worked less at $t = 1$ and has a high probability of solving the problem alone. Thus he rather collects the reputation gain r than to seek advice.

Proposition 1.4.
There exists a Perfect Bayesian Equilibrium with the following properties.

1. *Advice-seeking harms reputation, i.e.* $\hat{\mu}^{N,y_1} > \hat{\mu}^{A,y_2} = 0 \quad \forall y_1, y_2 \in \{S, F\}$.
2. *The agent receives positive reputation utility only if she does not ask for advice, i.e.* $\overline{\mu} \in \left(0, \hat{\mu}^{N,S}\right]$.
3. *Reputational benefits are intermediate, i.e.* $r \in \left(\frac{f(\lambda_h)}{2c_e}, \frac{f(\lambda_l)}{2c_e}\right]$.
4. *Reputation concerns discourage the high type from seeking advice: both types seek advice without reputation concerns, but only the low type seeks advice with reputation concerns, i.e.* $f(\lambda) > 0 \ \forall \lambda \in \{\lambda_l, \lambda_h\}$ *and* $\tilde{f}(\lambda_l) > 0 > \tilde{f}(\lambda_h)$.

Proof. To construct the equilibrium described in Proposition 1.4, I first note that in this separating equilibrium with reputation concerns only the low type seeks advice. Thus, advice-seeking is fully informative, i.e. $\hat{\mu}^A = 0$. It follows that $R^{A,S} = R^{A,F} = 0$ and $d^A = 0$. The respective types' effort provision in that equilibrium is given by $e_h^N = \frac{(1-\pi\lambda_h^2)(v+d^N)}{c_e}$ and $e_l^A = \frac{v+(1-\pi\lambda_l)d^N}{c_e}$.

I will show that there exists a PBE in which $d^N = 0$. In this case, $e_l^A > e_h^N$ and thus $\hat{\mu}^{N,S} < \hat{\mu}^{N,F}$. $d^N = 0$ then holds only if $\overline{\mu} \leq \hat{\mu}^{N,S}$. Therefore, no advice-seeking comes with reputation benefits: $R^{N,S} = R^{N,F} = r$. The condition for a separating equilibrium is then follows from Condition II: $(2 - \pi\lambda_l^2)\lambda_l v^2 - 2c_e r \geq 2c_e c_A \geq (2 - \pi\lambda_h^2)\lambda_h v^2 - 2c_e r$ which can be re-written as $f(\lambda_l) \geq 2c_e r > f(\lambda_h)$. As types seek advice without reputation concerns, it must also hold that $f(\lambda) \geq 0 \ \forall \lambda$. Therefore, the necessary conditions for the equilibrium to hold are (1) $f(\lambda_l) > f(\lambda_h) \geq 0$, (2) $r \in (\frac{f(\lambda_h)}{2c_e}, \frac{f(\lambda_l)}{2c_e}]$, (3) $\overline{\mu} \in (0, \hat{\mu}^{N,S}]$. □

In the equilibrium described by Proposition 1.4, reputation concerns discourage the high type from seeking advice. The first two conditions state that reputation utility is independent of project outcome and is positive if and only if the agent does not seek advice. In equilibrium, advice-seeking is a perfect signal for a low type since they have a higher incentive to seek advice (as they need advice more likely and work harder). Therefore, the high type fears for his reputation if seeking advice. If reputation concerns are sufficiently strong she is not willing to seek advice to keep her reputation. Thus there is a lower bound on r. In addition, if reputation concerns are too strong even the low type does not seek advice and deviates. This gives an upper bound on r.

If reputation concerns increase above that upper bound, they can even prevent the low type from seeking advice and thus result in no advice-seeking at all. In that case no type seeks advice and any off-equilibrium deviation to advice-seeking is punished by the neutral observer with a low posterior belief. Thus, with sufficiently strong reputation concerns no agent seeks advice. Proposition 1.5 shows that strong reputation concerns can always destroy any advice-seeking, regardless of who seeks advice without reputation concerns.

Proposition 1.5.
There exists a class of Perfect Bayesian equilibria with the following properties.

1. *Advice-seeking harms reputation, i.e.* $\hat{\mu}^{N,y_1} > \hat{\mu}^{A,y_2} \;\; \forall y_1, y_2 \in \{S, F\}$.
2. *The agent receives positive reputation utility only if she does not ask for advice, i.e.* $\overline{\mu} \in \left(\max\left\{\hat{\mu}^{A,S}, \hat{\mu}^{A,F}\right\}, \hat{\mu}^{N,F}\right]$.
3. *Reputational benefits are sufficiently large, i.e.* $r > \max\left\{\frac{f(\lambda_h)}{2c_e}, \frac{f(\lambda_l)}{2c_e}\right\}$.
4. *Reputation concerns destroy all advice-seeking: at least one type seeks advice without reputation concerns, but no type seeks advice with reputation concerns, i.e.* $(f(\lambda_l) > 0 \;\; \vee \;\; f(\lambda_h) > 0)$ *and* $\tilde{f}(\lambda) < 0 \;\; \forall \lambda \in \{\lambda_l, \lambda_h\}$.

Proof. In all equilibria described in Proposition 1.5, no type seeks advice with reputation concerns. I will construct equilibria in which it holds that $d^A = d^N = 0$. In this case, $e_h^N = \frac{(1-\pi\lambda_h^2)v}{c_e} > \frac{(1-\pi\lambda_l^2)v}{c_e} = e_l^N$. Hence, success is an indicator for being a high type, conditional on no advice-seeking. Thus, $\hat{\mu}^{N,S} > \hat{\mu}^{N,F}$. To construct the equilibria, suppose that it holds for the off-equilibrium beliefs (i.e. if the agent sought advice) that $\hat{\mu}^{A,y} < \hat{\mu}^{N,F}\ \forall y$. Then advice-seeking in punished in terms of reputation utility if $\overline{\mu} \in \left(\max\left\{\hat{\mu}^{A,S}, \hat{\mu}^{A,F}\right\}, \hat{\mu}^{N,F}\right]$, and thus $R^A = 0$ and $R^N = r$.

There are three cases consistent with Proposition 1.5: (a) $f(\lambda_l) > f(\lambda_h) \geq 0$, (b) $f(\lambda_l) \geq 0 > f(\lambda_h)$, (c) $f(\lambda_h) \geq 0 > f(\lambda_l)$. Furthermore, I can re-write Condition II for the pooling equilibrium with $d^A = d^N = 0$ as $(2 - \pi\lambda^2)\lambda v^2 - 2c_e r - 2c_e c_A < 0$, i.e. $f(\lambda) - 2c_e r < 0\ \forall\lambda$. Therefore, any PBE described by the following conditions are consistent with Proposition 1.5:
(1) $\max\{f(\lambda_l), f(\lambda_h)\} \geq 0$, (2) $r > \max\left\{\frac{f(\lambda_l)}{2c_e}, \frac{f(\lambda_h)}{2c_e}\right\}$, (3) $\hat{\mu}^{A,y} < \hat{\mu}^{N,F} < \hat{\mu}^{N,S}\ \forall y$, (4) $\overline{\mu} \in \left(\max\left\{\hat{\mu}^{A,S}, \hat{\mu}^{A,F}\right\}, \hat{\mu}^{N,F}\right]$. □

This completes the section on the negative effects of reputation concerns. It shows that the model captures previous findings by Levy (2004) and Chandrasekhar et al. (2018). If the agent's evaluation only depends on his decision to seek advice and advice-seeking signals a low type, as in Propositions 1.4 and 1.5, agents fear for their reputation and do not seek advice. However, the next section shows opposite effects; in fact there exist equilibria in which (a) reputation concerns increase advice-seeking and (b) advice-seeking even signals a high type.

On the Positive Effects of Reputation Concerns

This section reveals two distinct channels how reputation concerns can increase advice-seeking. Theorem 1.1 shows that reputation concerns can raise effort provision which increases the incentives to seek advice. Indeed, in the equilibrium described therein reputation concerns induce both types to seek advice. The intuition is as follows. Suppose without reputation concerns advice-seeking costs are too high so that no type seeks advice. In the equilibrium with reputation concerns, the agent only receives positive reputation utility if her project is successful. Therefore the expected benefits of a successful project increase and both types exert high effort. Higher effort implies higher incentives to seek advice. If reputation concerns are strong, effort provision and hence the incentives to seek advice increase and outweigh the high costs of advice-seeking. Consequently both types seek advice.

Theorem 1.1.
There exists a Perfect Bayesian Equilibrium with the following properties.

1. *Project success increases reputation, i.e.* $\hat{\mu}^{a,S} > \hat{\mu}^{a,F} \;\; \forall a \in \{A, N\}$.

2. *The agent receives positive reputation utility only if her project is successful and she does not ask for advice, i.e.* $\overline{\mu} \in \left(\max\left\{\hat{\mu}^{A,S}, \hat{\mu}^{N,F}\right\}, \hat{\mu}^{N,S}\right]$.

3. *Reputational benefits are intermediate, i.e.* $r \in [\underline{r}, \overline{r}]$*, and bounds are given in the proof.*

4. *Reputation concerns induce both types to seek advice: no type seeks advice without reputation concerns, but both types seek advice with reputation concerns, i.e.* $f(\lambda) < 0 \;\; \wedge \;\; \tilde{f}(\lambda) \geq 0 \;\; \forall \lambda \in \{\lambda_l, \lambda_h\}$.

Proof. The equilibrium described by Theorem 1.1 states that no type seeks advice without reputation concerns but all types seek advice with reputation concerns. I construct an equilibrium in which $d^N = r$ and $d^A = 0$ and both types seek advice under reputation concerns. Then, $e_h^A = \frac{v+(1-\pi\lambda_h)r}{c_e} > \frac{v+(1-\pi\lambda_l)r}{c_e} = e_l^A$. As both types seek advice and the low type is more likely to need advice, it follows that $\hat{\mu}^{N,S} > \hat{\mu}^{A,S}$ and $\hat{\mu}^{N,F} > \hat{\mu}^{A,F}$. The effort levels imply that $\hat{\mu}^{N,S} > \hat{\mu}^{N,F}$ and $\hat{\mu}^{A,S} > \hat{\mu}^{A,F}$. Taken together, the following ordering of posterior beliefs arises: $\hat{\mu}^{A,F} < \min\left\{\hat{\mu}^{A,S}, \hat{\mu}^{N,F}\right\} < \max\left\{\hat{\mu}^{A,S}, \hat{\mu}^{N,F}\right\} < \hat{\mu}^{N,S}$. If $\overline{\mu} \in \left(\left\{\hat{\mu}^{A,S}, \hat{\mu}^{N,F}\right\}, \hat{\mu}^{N,S}\right]$, it holds that $R^{A,F} = R^{A,S} = R^{N,F} = 0$ and $R^{N,S} = r$. It follows that $d^N = r$ and $d^A = 0$ hold. For both types to seek advice in equilibrium, from (Condition II) it follows that

$$\begin{aligned} &[(2-\pi\lambda^2)(v+r) - \pi\lambda r][\lambda v - (1-\lambda)r] \geq 2c_e c_A \;\; \forall \;\; \lambda \in \{\lambda_l, \lambda_h\} \\ &\Leftrightarrow \;\; f(\lambda) - r(z \cdot r - w \cdot v) \geq 0 \qquad \forall \;\; \lambda \in \{\lambda_l, \lambda_h\} \end{aligned} \tag{1.20}$$

with $w = 2((2-\pi\lambda^2)\lambda - 1)$ and $z = (1-\lambda)(2-\pi\lambda-\pi\lambda^2) > 0$ must hold. First, note that the no advice-seeking pooling equilibrium without reputation concerns implies $f(\lambda) < 0 \, \forall \lambda \in \{\lambda_l, \lambda_h\}$. Hence, reputation concerns increase advice-seeking only if $zr - wv < 0$, i.e. $r < \frac{wv}{z}$ with $w > 0$. Secondly, as $z > 0$, the left-hand side of (1.20) is concave in r with a maximum at $\frac{wv}{2z}$. The zeros of Equation (1.20) are given by $r = \frac{wv \pm \sqrt{(wv)^2 + 4zf(\lambda)}}{2z} \in (0, \frac{wv}{z})$ as $f(\lambda) < 0$. Therefore, I can construct a PBE, described in Theorem 1.1, with the following properties:

1. $\hat{\mu}^{A,F} < \min\left\{\hat{\mu}^{A,S}, \hat{\mu}^{N,F}\right\} < \max\left\{\hat{\mu}^{A,S}, \hat{\mu}^{N,F}\right\} < \hat{\mu}^{N,S}$,

2. $\overline{\mu} \in \left(\max\left\{\hat{\mu}^{A,S}, \hat{\mu}^{N,F}\right\}, \hat{\mu}^{N,S}\right]$,

3. $r \in [\underline{r}, \overline{r}] \;\; \forall \lambda \in \{\lambda_l, \lambda_h\}$, with $\underline{r} = \frac{wv - \sqrt{(wv)^2 + 4zf(\lambda)}}{2z}$ and $\overline{r} = \frac{wv + \sqrt{(wv)^2 + 4zf(\lambda)}}{2z}$,

4. $w = 2((2 - \pi\lambda^2)\lambda - 1) > 0 \ \forall \lambda \in \{\lambda_l, \lambda_h\}$,

5. $-\frac{(wv)^2}{4z} < f(\lambda) < 0 \ \wedge \ \tilde{f}(\lambda) \geq 0 \ \forall \lambda \in \{\lambda_l, \lambda_h\}$. □

In equilibrium, both type seek advice but the high type exerts higher effort ($e_h^A = \frac{v+(1-\pi\lambda_h)r}{c_e} > \frac{v+(1-\pi\lambda_l)r}{c_e} = e_l^A$). Consequently, project success is associated with high ability and thus increases reputation, as stated in the first condition. Yet advice-seeking itself still signals a low type. In fact, both types seek advice *despite* a subsequent loss of reputation. At the advice stage reputation concerns induce a trade-off between reputation and expected benefits from a successful project. First, reputation concerns increase the incentive to not seek advice as only then the agent may collect the gains from reputation. Secondly, they raise effort provision which in turn increases the expected benefits from advice-seeking. In Theorem 1.1, the latter effect is stronger. Thus, reputation concerns increase the incentives to seek advice.

Moreover, reputation concerns are bounded. On the one hand, r must be sufficiently large to induce both types to exert sufficient effort which induces them to seek advice in the first place. Yet, because advice-seeking is also associated with a loss of reputation, r must be bounded from above. Otherwise the high type has an incentive to deviate, work hard but not seek advice. In that case, she still has a sufficiently high probability of solving the problem alone *and* receive r.

Despite the positive effect of reputation concerns on actual advice-seeking, any agent who seeks advice must still fear for their reputation in Theorem 1.1. This is different in Theorem 1.2. Here the agent's reputation *increases* with advice-seeking because only the high type seeks advice. The underlying intuition is the following. Suppose the high type gains more from advice-seeking than the low type, similar to the intermediate sorting equilibrium in Section 1.3. Yet advice-seeking costs are high and no type seeks advice without reputation concerns. In equilibrium, reputation concerns induce only the high type to seek advice; hence advice-seeking becomes a perfect signal for high ability. Thus, the high type receives additional reputation utility that induces him to seek advice. For the low type, reputation concerns are not high enough; she abstains from advice-seeking and a separating equilibrium emerges.

Theorem 1.2.
There exists a Perfect Bayesian Equilibrium with the following properties.

1. *Advice-seeking increases reputation, i.e.* $\hat{\mu}^{A,y_1} = 1 > \hat{\mu}^{N,y_2} \quad \forall y_1, y_2 \in \{S, F\}$.

2. *The agent receives positive reputation utility only if she asks for advice, i.e.* $\overline{\mu} \in \left(\hat{\mu}^{N,y}, \hat{\mu}^{A,y}\right] \quad \forall y \in \{S, F\}$.

3. *Reputational benefits are intermediate, i.e.* $r \in \left[\frac{-f(\lambda_h)}{2c_e}, \frac{-f(\lambda_h)}{2c_e}\right)$.

4. *Reputation concerns induces only the high type to seek advice: no type seeks advice without reputation concerns, but the high type seeks advice with reputation concerns, i.e.* $f(\lambda) < 0 \;\; \forall \lambda \in \{\lambda_l, \lambda_h\} \quad \wedge \quad \tilde{f}(\lambda_h) \geq 0 > \tilde{f}(\lambda_l)$.

Proof. In the PBE described by Theorem 1.2, the types separate under reputation concerns as the high type seeks advice while the low type does not. Thus, advice-seeking is perfectly informative and $\hat{\mu}^{A,S} = \hat{\mu}^{A,F} = 1$ and consequently $R^{A,S} = R^{A,F} = 1$ and $d^A = 0$. I construct an equilibrium in which it holds that $d^N = 0$. In this case, effort is given by $e_h^A = \frac{v}{c_e}$ and $e_l^N = \frac{(1-\pi\lambda_l^2)v}{c_e}$. It follows that, conditional on no advice-seeking, success is an indicator for the high type and thus $\hat{\mu}^{N,S} > \hat{\mu}^{N,F}$. To sustain $d^N = 0$, I impose $\overline{\mu} \in (\hat{\mu}^{N,S}, 1]$.

Therefore, Condition II implies for this specific PBE in which both types seek advice that $(2 - \pi\lambda_h^2)\lambda_h v^2 + 2c_e r \geq 2c_e c_A > (2 - \pi\lambda_l^2)\lambda_l v^2 + 2c_e r$. This can be re-written as $f(\lambda_h) > -2c_e r > f(\lambda_l)$. It follows that there exists a Perfect Bayesian Equilibrium, described by Theorem 1.2, with the following conditions:
(1) $r \in \left[\frac{-f(\lambda_h)}{2c_e}, \frac{-f(\lambda_l)}{2c_e}\right)$, (2) $\overline{\mu} \in (\hat{\mu}^{N,S}, 1]$, (3) $f(\lambda_l) < f(\lambda_h) < 0$. □

There is one common theme that combines all equilibria with a positive effect of reputation concerns. They are all based on the effects of effort on advice-seeking. For the equilibrium described in Theorem 1.1, reputation concerns affect effort provision, change the likelihood of project success and thus indirectly influence the decision to seek advice. The equilibrium described in Theorem 1.2 is based on the presumption that the high type has higher benefits from advice-seeking without reputation concerns. Otherwise, reputation concerns could not separate low and high types. But the positive effect of ability on advice again builds on the complementarity between effort and advice; hence the equilibrium in Theorem 1.2 does not exist without effort considerations.

Therefore, the positive results on reputation concerns hinge on the incorporation of effort into the model. If that part is excluded from the analysis, there is no positive effect of ability on advice; and reputation concerns have an unambiguously negative effect on advice-seeking. Consequently, models that do not take into account the joint interplay between effort, advice, and reputation miss potential positive effects of reputation concerns. This was the case in previous advice-seeking models such as Levy (2004) and Chandrasekhar et al. (2018) who conclude that advice-seeking signals low ability and thus reputation concerns reduce advice-seeking. Theorem 1.2 stands in stark contrast to that finding. Because ability increases effort provision competent agents have higher incentives to seek advice. Ultimately, a separating equilibrium emerges in which agents should not fear for their reputation, but in fact seek advice precisely because *advice-seeking signals high ability*.

1.5 Discussion

The key new feature of this model is the introduction of effort into a model of advice-seeking. Such a setting not only mimics many relevant situations such as employees seeking help from their supervisors, also its implications relate well to existing empirical evidence. This section first discusses that evidence in relation to the model's findings as well as further implications and gives a thorough overview over potential new avenues of research.

1.5.1 On the Effects of Advice

Advice is helpful. It increases the chances to succeed in the task at hand. Thus, if success depends on the agent's effort provision, advice and effort become complements. If the availability of advice increases it is more valuable to work hard. Lemma 1.1 describes that complementarity. Effort provision and therefore (expected) performance increase with advice-seeking and the availability of advice. Due to the prevalent unobservability of effort, empirical research has focused on the effect of advice on performance and finds that, unsurprisingly and in line with the model, advice increases performance in various settings. For a detailed and excellent overview on that finding, I refer the reader to Bonaccio and Dalal (2006).

Moreover, empirical evidence shows that perceived support increases an employee's performance. In terms of the model, perceived support can be interpreted a proxy for the (perceived) availability of advice. Mueller (2012) shows that larger team sizes result in lower perceived support which in turn decreases performance in larger teams. Relatedly, Amabile et al. (2004) find that higher perceived leader support increases employees' creative work performance. This result also relates to current research on the value and role of supervisors. Lazear et al. (2015) estimate that replacing a bad with a good supervisor is similar to adding one employee to a nine-member team. As an explanation to these large productivity effects of supervisors, Lazear et al. (2015, p. 585) cite from interviews with managers and workers who "all emphasized the significant effects that bosses have in coaching and motivating workers." That fits the model's interpretation of supervisors as advisors. Bosses coach their employees by providing advice when needed and that gives them an additional motivation.

Further, advice availability increases advice-seeking due its positive effects on effort provision. There is ample evidence for this positive relationship across various fields. In the medical sector, Hofmann et al. (2009) show that nurses seek more help when the helper is perceived as more accessible or is formally required to help. In education, students seek more help and advice if their teachers are more approachable (Ryan et al. 2001; Johnson and LaBelle 2017). In organizations, Borgatti and Cross (2003) and van der Rijt et al. (2013) find that employees' seek more advice from more available advisors.

These findings have manifold implications for corporate culture and the role of leaders. In recent years, many firms have installed policies to foster communication and collaboration in order to adapt to a ever-changing world. More and more jobs involve non-routine tasks (Autor et al. 2003) and social skills (Deming 2017; Deming and Kahn 2018). At the same time, hierarchies get flatter (Rajan and Wulf 2006), companies implement open-door policies, open plan offices and social intranets and emphasize the need for collaboration in employee handbooks and their corporate statements.[8] In terms of the

[8]For example, HP inc. has installed an open-door policy "for a work environment" where "open, honest communication between managers and employees is a day-to-day business practice" and "advice is freely given", see `hp.com/hpinfo/abouthp/diversity/open-door` (visited on 2019-09-06). Facebook's new headquarters includes an open plan office for around 2,800 employees to "make work as frictionless as possible", accord-

model, these policies are similar to a decrease in advice-seeking costs or an increase of advice availability. In that case they will not only increase advice-seeking but also motivate employees to work hard which gives an additional positive effects of advice on performance.

1.5.2 On Advice, Ability, and Reputation

Starting with Tessler and Schwartz (1972) and Ashford and Cummings (1983) psychological and organizational research on help-, advice- and feedback-seeking has been concerned with the seeker's image. In fact, DePaulo and Fisher (1980) find in an early lab experiment on help-seeking that subjects think they appear less competent when seeking help. Further evidence from different fields and environments is provided and reviewed by Morrison and Bies (1991), Ashford and Northcraft (1992), Lee (1997), Ryan et al. (2001), van der Rijt et al. (2013), and Yoon et al. (2019). The main finding is that individuals fear for their reputation if they seek advice, help, information or feedback and that such fear makes individuals seek advice less.[9]

The underlying intuition was formalized and introduced to economics by Levy (2004). In her model, an image-concerned decision maker needs to take an action in an uncertain environment. Before choosing the action, he receives a private signal about the state of the world. The signal's precision increases with the decision maker's ability. In addition he can (publicly) consult an advisor to receive an second signal. Levy (2004) shows that decision makers with high ability do not consult the advisor if image concerns are sufficiently strong. As signal precision increases with ability, a decision maker that is not seeking advice signals a high own signal precision and thus high ability. On the other hand, low types need the signal provided by the advisor to take an informed decision; hence they seek advice despite reputation losses. A separating equilibrium emerges. Levy (2004) also find that this equilibrium is

ing to the Washington Post (`wapo.st/1NYmOWF`, visited on 2019-09-06). The Daimler AG has implemented a new Enterprise Social Network for its 300,000 employees in 2018, see `blog.daimler.com/en/2018/10/29/digital-life-daimler-collaborate-networking` (visited on 2019-09-06). Amabile et al. (2014) give an example of a "corporate culture of helping".

[9] Such fear is also discussed in newspaper articles, for instance in "Why Is Asking for Help So Difficult?", New York Times, July 05, 2007, p. C5.

inefficient; decision makers who should seek advice from a welfare perspective end up not seeking advice.

Chandrasekhar et al. (2018) investigate the effects of signaling concerns and shame on advice-seeking.[10] They first set up a model that is similar to Levy (2004) in spirit and results; they show that if low competence implies higher benefits from advice, reputation concerns decrease advice-seeking. Moreover, they run a lab-in-the field experiment in India to show further proof of negative effects of signaling and shame on advice-seeking. Chandrasekhar et al. (2018) first find that subjects with high ability do not seek advice since they regard it as unnecessary. Further, the advice-seeking probability of subjects with low ability decreases by 55% when the need for advice becomes correlated with ability. Thus, subjects care for their reputation and fear signaling incompetence when seeking advice.

The intuition that advice-seeking signals inability is captured in the current model by the Advice Effect. Individuals with low ability have a lower probability of solving a problem on their own. Hence advice becomes more beneficial for them and as a consequence advice-seeking signals a low type. In Levy (2004), this effect is implemented as a difference in the precision of private signal; advice is more beneficial to decision makers with low ability and imprecise signals. Chandrasekhar et al. (2018) plainly assume that low ability types have higher benefits from advice.

The current model adds an additional layer to these findings. Introducing effort gives rise to a positive effect of ability on advice via effort provision. Since ability increase effort and advice and effort are complements, ability increases the incentives to seek advice. This Effort Effect implies an counteracting channel compared to the Advice Effect. In such richer setting, the agent can in fact signal high ability by seeking advice in some situations.

While this theoretical result contrasts the aforementioned literature, some experimental evidence is consistent with such positive relationship between reputation and advice-seeking. Brooks et al. (2015) investigate how subjects perceive the competence of advice-seekers in a series of lab experiments. They find that the perceived competence increases if subjects are asked for their

[10]Chandrasekhar et al. (2018) model shame as disutility of incompetent agents from interacting with someone who knows of their incompetence and show that it is an important driver of advice-seeking; it decreases the advice-seeking probability of subjects with low ability by 65%.

advice, in particular for difficult tasks and independent of prior information about the advice-seeker's performance. In a related study, Thompson and Bolino (2018) focus on negative beliefs about accepting help from co-workers. They find that employees who hold those beliefs and, for example, think that "[m]y manager believes in me more when I complete my work without help" are evaluated worse in a variety of dimensions such as performance, competence and creativity. Hence positive reputation effects of accepting help seem to exist. Similarly, Yoon et al. (2019) find that asking for time extensions of project work leads to better evaluations by the supervisor. Employees, however, do not expect such updating but in fact fear appearing incompetent when asking for extensions. Again, this fear prevents them from asking for more time when they need it. More generally, Brooks and John (2018) or Grant (2018) argue that people should ask more questions and ask for more help because, for instance, help-seeking increases the seeker's likability.

1.5.3 Open Questions

The inconclusive evidence on the link between reputation and advice-seeking hints towards a more general point. Advice-seeking is a complex decision that is highly dependent on situational, personal and organizational factors. Therefore, general claims are hard to make and the exact interaction of advice and reputation depends strongly on the circumstances. That presents a fruitful avenue for economists who build models and analyze data that are often tied to one specific environment instead of working on one general, ubiquitously applicable theory (Rodrik 2015). Further open questions on advice-seeking and related matter are discussed in the following.

The Relationship between Advice and Effort

The model's findings build strongly on a complementarity between ability and effort. Yet one can easily imagine situations in which the relationship is reversed. For example, team meetings are institutionalized events of advice-seeking and -giving. They are central to organizations, but yet they are criticized for being inefficient and time-wasting.[11] An exemplary anecdote is

[11] See Rogelberg et al. (2007) or Perlow et al. (2017). A Microsoft survey in 2015 finds that employees spend roughly 5.5 hours per week in meetings, and seven of ten employees

given by Carnegie (1948) who describes his publisher Leon Shimkin's frustration with the inefficiency and unnecessary length of team meetings. But Shimkin realized that his employees were using team meetings, i.e. advice, as a substitute for problem-solving effort. They overloaded meetings with their own problems which made an efficient use of the meetings impossible.

This anecdote reveals the relationship between advice and effort is highly dependent on the kind of work under examination. The current model focuses on effort that determines expected benefits from the employee's work. On the other hand, an employee's work also consists of problem-solving effort, especially in complex, creative and non-routine tasks. Here, advice and problem-solving effort are substitutes. In that case, that advice-seeking and the availability of advice decrease (problem-solving) effort. The intuition is straightforward. If an agent knows that she will receive problem-solving advice in the future, it is not worthwhile to bear own problem-solving effort costs. In this case, the principal may want to restrict advice-seeking incentives as Leon Shimkin did. To shorten the length of team meetings, he required his employees to fill out a pre-meeting questionnaire if they wanted to discuss a problem and its potential solution. By doing so, he made them think about their problem before the meeting and therefore (a) increased advice-seeking costs and (b) induced them to increase own problem-solving effort.

Optimal Firm Policies

Like Leon Shimkin, firms implement policies to create an efficient workspace. As jobs have changed dramatically towards complex, non-routine jobs that require much teamwork (Autor et al. 2003; Autor and Price 2013; Bandiera et al. 2013; Deming 2017), firms implement open-door policies, open plan workspaces and enterprise social networks to increase accessibility and decrease communication costs (see footnote 8). Or they create a corporate culture of helping to build a norm of helping and foster cooperation and teamwork (Cleavenger et al. 2007; Amabile et al. 2014; Grodal et al. 2015).

On the other hand, well-intentioned corporate policies to foster collaboration and prosociality can turn out inefficient or even profit-harming. Haas and Hansen (2005) and Hansen (2009) describe cases where firms' emphasis

find them ineffective, see `news.microsoft.com/2005/03/15/survey-finds-workers-average-only-three-productive-days-per-week/` (visited on 2019-09-06).

on teamwork and cooperation led to inefficient because unnecessary collaboration. Even "organizational citizen behavior" (e.g. helping others, speaking up, taking on responsibilities) can harm individual and organizational performance (Bolino et al. 2013; Bolino and Grant 2016) and "helping routines" in organizations can result in inefficient help-seeking and a crowding-out of problem-solving effort (Hargadon and Bechky 2006, p.490).

These examples call for a thorough examination of corporate policy effects. In the current model there are no costs associated to the availability or provision of advice; thus advice is always beneficial. However, a potential substitutability between advice and problem-solving effort as well as opportunity costs of advising yield natural upper bounds to optimal advice-seeking. Consequently, firms may need to restrict advice-seeking by making managers unavailable (e.g. by increasing advisors' workload, managers' span of control, by changing management practices, and consequently lowering α) or increasing the costs of advice-seeking c_A. A complete analysis of optimal corporate policies with regard to advice-seeking incentives is left to future research.

Behavioral Aspects of Advice-Seeking

The willingness to seek advice negatively depends on the perceived probability of receiving advice (in the model captured by α). If an individual thinks she will not get advice she does not seek it. However, there are strong misperceptions regarding advice-seeking and -giving behavior. Research in psychology (with a focus on helping) shows that individuals *underestimate* the willingness to help as well as the amount of help they would receive if asking for help (see Flynn and Bohns (2008), Newark et al. (2017) and the references therein). On the other side of the interaction, potential helpers *overestimate* the amount of help-seeking (Bohns and Flynn 2010) because they do not account for a feeling of "embarrassment" when seeking help. This provides an interesting link to the findings by Brooks et al. (2015). In both cases, helpers or advisors do not believe that negative emotions (such as feeling incompetent or embarrassed) play a role in helping contexts but they do. Therefore, the case of advice- or help-seeking constitutes an important application of economic research on misperceptions of own and others' attitudes and preferences (see Eckel and Grossman 2008; Fedyk 2018; Ericson and Laibson 2019).

1.6 Conclusion

There is a common belief among researchers (Lee 2002) as well as students (Ryan et al. 2001), employees (Thompson and Bolino 2018) and journalists[12] that advice-seeking harms the reputation of the seeker. Consequently, reputation concerns also harm advice-seeking incentives and lead to inefficient outcomes (Levy 2004). This holds true if lower ability implies a higher need for advice and consequently higher incentives to eventually seek it. However, recent evidence shows that advice-seeking can also have positive implications for the seeker's reputation (Brooks et al. 2015).

Consistent with these findings, the current paper has revealed the belief of reputation losses from advice-seeking to be partly flawed as its intuition is simplistic. If advice-seeking is embedded into a richer environment, incompetent agents may not seek advice and reputation concerns have positive effects on advice-seeking. Because ability increases effort provision agents with higher ability work harder. Therefore, they have more at stake when needing advice and thus have a higher incentive to seek advice. The same intuition explains why reputation concerns can induce agents to seek advice, via two distinct mechanisms. First, if project success is associated with high ability, any agent exerts higher effort and thus the incentive to seek advice increase for all types. Secondly, in some situations only agents with high ability seek advice. In that case, advice-seeking itself signals high ability.

Many unanswered questions surround the decision to seek advice that are (partly) discussed in Section 1.5.3. For example, should a company issue policies that enhance or discourage advice-seeking? When does advice increase effort, and when does it lead to a crowding-out? What are the reasons for common misperceptions associated with advice-seeking? Other questions involve the advisor's incentives. Absent formal requirement, why do people help each other? How can an organization make sure that seeking advice and advising is an efficient process? Given the manifold circumstances, in which advice and help are important factors, there is opulent work to be done. This model is to be thought of as a first step towards a more thorough analysis of advice-seeking in various situations.

[12] "Why Is Asking for Help So Difficult?", New York Times, July 05, 2007, p. C5.

Chapter 2

Optimal Rule Design and an Emerging Moral Wiggle Room

2.1 Introduction

Individuals behave more selfishly when their actions' observability decreases. In dictator games, dictators give less when it is uncertain whether unfair outcomes result from their selfish behavior or from bad luck (Andreoni and Bernheim 2009), a lack of information, or other dictators' behavior (Dana et al. 2007). They even avoid information to uphold an excuse to act selfishly (Dana et al. 2007; Grossman 2014). Such "moral wiggle room" behavior can be explained by selfish, but image-concerned individuals (Andreoni and Bernheim 2009). These individuals want to be perceived as good people, thus as being fair in dictator games; yet they also want to maximize their own monetary payoff. When their actions' observability decreases, it is possible to act selfishly and uphold a positive image at the same time, for instance by blaming bad luck for unfair outcomes.

In this chapter, I investigate how the optimal design of non-binding rules, issued to promote prosocial behavior, is shaped by selfish behavior in moral wiggle rooms that emerge under some rules. Rules are omnipresent; they shape all aspects of life, public or private. Governments enact laws; parents set rules for their children; religious leaders create commandments to follow. Given the widespread use of rules, it is unsurprising that they differ widely in content and scope. For example, many rules are broad and universal ("You shall not lie"), some rely on the individual's judgment ("Do the right thing") and others are highly specific and conditional on the situation ("If A, then do B").[1]

I argue that a moral wiggle room emerges under highly specific rules, but not under universal rules. Using both a model and an experiment, I compare behavior and welfare consequences under two rules that are designed to induce an agent to act prosocially. The "Unconditional Rule" prescribes one single action regardless of circumstances; it is thus universal. The "Conditional Rule" prescribes different actions conditional on the state of the world; it is thus specific. If all agents followed rules blindly, the Conditional Rule would be superior because of its conditionality. It can prescribe the efficient action for each state of the world. To the contrary, the Unconditional Rule dictates inefficient actions in some situations.

The model clarifies that only the Conditional Rule creates a moral wiggle room, due to the existence of selfish, but image-concerned agents. Consequently it induces less compliance, less prosocial actions and thus lower welfare. The intuition is the following. The two goals of selfish, image-concerned agents conflict under the Unconditional Rule that prescribes one single prosocial action. Here, acting selfishly is a perfect signal for being selfish. If image concerns are sufficiently strong those agents will follow the rule to prevent appearing selfish. On the other hand, the Conditional Rule prescribes selfish behavior in some situations (i.e. when it is efficient to act selfishly). In an uncertain world, a neutral observer thus can never be certain about an agent's intentions. Either the agent acted selfishly on purpose, or he is a rule-follower that received a misleading signal about the state of the world. Hence selfish, image-concerned agents can maintain a positive image even when acting selfishly, due to the emerging moral wiggle room.

Therefore, while being inefficient in some situations the Unconditional Rule also leads to more rule-following than the Conditional Rule. If the latter part is of particular value, the Unconditional Rule is superior. In the theory part of this chapter, I first find conditions for the existence and uniqueness of a wiggle room equilibrium and then compare welfare implications of both rules. The wiggle room can in fact make the Conditional Rule inferior to the

[1] Anyone working in an university should be aware of truly specific rules, for example for travel reimbursements.

Unconditional Rule, if there are sufficiently many selfish, image-concerned agents and image concerns are sufficiently pronounced.

In the second part of the chapter, I present results from a laboratory experiment to test for the existence of a moral wiggle room under the Conditional Rule. For that purpose, I investigate how subjects react to different rules in a variant of the public goods game. Subjects contribute to a public good and are ought to follow a non-binding rule. At the same time, uncertainty blurs the relationship between the subjects' actions and actual consequences. In two treatments, I examine the effects of the Unconditional Rule and the Conditional Rule on contribution behavior.

The experimental findings can be summarized as follows. (1) Subjects tend to follow both rules. Their contributions are strongly affected by what the rule prescribes them to do. (2) There is scope for wiggle room behavior as it is socially less inappropriate to act selfishly (i.e. contribute zero to the public good) under the Conditional Rule. (3) In general, the wiggle room has no effect on overall contributions. When the Conditional Rule and the Unconditional Rule prescribe the same behavior, the wiggle room equilibrium predicts *less* adherence to the Conditional Rule. Yet, contributions in the experiments (mostly) do not differ in the hypothesized way. (4) Nonetheless, the results also indicate that selfish subjects do react to the potential moral wiggle room under the Conditional Rule when incentives to act selfishly are sufficiently strong.

The contribution of this chapter is therefore two-fold. First, the theory clarifies the welfare effects of rules. I show under what conditions a wiggle room equilibrium exists that makes the Unconditional Rule worthwhile to use. In that case, the Conditional Rule allows for excuses of selfish actions and are thus induces less rule-following. The experimental evidence on wiggle room behavior under the two different rules remains, however, inconclusive. Even though selfish subjects follow the Conditional Rule less when incentives are sufficiently strong, there is no pronounced general effect of the emerging moral wiggle room.

Related Literature. This chapter is related to two different strands of literature. First, the literature on moral wiggle rooms shows that individuals use ambiguity about their actions to act more selfishly. In dictator games, dictators give less when they can attribute their selfish behavior to bad luck (Dana et al. 2007; Andreoni and Bernheim 2009) or to other subjects' behavior (Dana et al. 2007). They even avoid information to uphold the moral wiggle room and thus their excuse to act selfishly (Dana et al. 2007; Grossman 2014). On the other hand, van der Weele et al. (2014) do not find wiggle room behavior in trust and moonlighting games. They explain that result by strong effects of reciprocity that outweigh the incentive to act selfishly in a moral wiggle room.

In general, wiggle room behavior can be explained by selfish agents who want to be perceived as "good" people.[2] Similar to this work, Andreoni and Bernheim (2009) theoretically and experimentally show that (a) image concerns of otherwise selfish agents can account for prosocial behavior in dictator games, (b) a norm of equal split emerges, and (c) that this norm breaks down when the dictators' true action is obscured to the receiver.[3] However, the focus of Andreoni and Bernheim (2009) differs from mine. They analyze behavior in dictator games only and thus are not able to examine welfare implications of the moral wiggle room. In contrast, my model compares prosocial behavior and welfare under the two different rules. I find that welfare increases under the Unconditional Rule only if externalities are high, costs of the prosocial action are low and the share of selfish agents is high (given the wiggle room equilibrium). Bénabou et al. (2018) find similar results when examining the impact of imperatives and narratives issued by a principal to affect an agent's prosocial behavior. In their setting, the principal has superior information about an action's externality and the question is how to effectively communicate his information to the agent. In one extension, imperatives bear a cost of flexibility, which is reminiscent of the Unconditional Rule's inefficiency. Similarly, Bénabou et al. (2018) find that the use of imperatives increases in the action's externality and decreases when the agent is less likely to follow the imperative.

On the experimental side, my work relates to laboratory experiments that investigate behavior under laws, obligations and advice. The results are in

[2] Whereas the interpretation of "good" varies by context. In dictator games, a good type is fair (Andreoni and Bernheim 2009), in the current context a good type follows rules.

[3] Thus, both their and my model build on the manifold work on the role of image concerns, on the theory side by Bénabou and Tirole (2006) and Ellingsen and Johannesson (2008) and on the empirical side see Bursztyn and Jensen (2017) for an excellent overview.

line with findings by Galbiati and Vertova (2008, 2014). In both papers, the authors find that minimum obligations increase contributions in public good games. Galbiati and Vertova (2014) also show that this is independent of formal sanctions for non-compliance. In contrast, Tyran and Feld (2006) find that only self-imposed obligations, so-called mild laws, affect subjects' contribution behavior. My experimental findings also relate to results on the effects of advice, see Croson and Marks (2001), Schotter and Sopher (2006), Chaudhuri et al. (2008) and Çelen et al. (2010). They show that advice, issued by other participants in the experiment, can also affect behavior and increase performance. As non-binding rules could also be interpreted as advice, the results from the current experiment, showing that subjects follow rules, fit well into that literature.

The chapter proceeds by presenting the model and its results on the wiggle room equilibrium as well as welfare implications of the two rules in Section 2.2. In Section 2.3, I present the experimental setup and results. Section 2.4 concludes.

2.2 Model

The model deals with the heterogeneous effects of the Unconditional Rule and the Conditional Rule on rule-following behavior. The analysis shows the existence and uniqueness of a "wiggle room equilibrium" under certain conditions. In this equilibrium, everyone follows the Unconditional Rule. The Conditional Rule, however, induces only partial rule-following as a selfish, image-concerned agent does not adhere to it.

2.2.1 Set-up

0. Overview There is one principal (she) and one agent (he). The principal issues a rule to induce the agent to take a prosocial action. The model consists of five stages, $t = 1, 2, 3, 4, 5$. At $t = 1$, the principal designs a non-binding rule to maximize expected welfare that is affected by the agent's action. At $t = 2$, Nature privately draws a state of the world ω. At $t = 3$, Nature draws a noisy signal about the realized state that is privately displayed to the agent. At $t = 4$, the agent chooses an action a, at a cost of $c(a)$, that

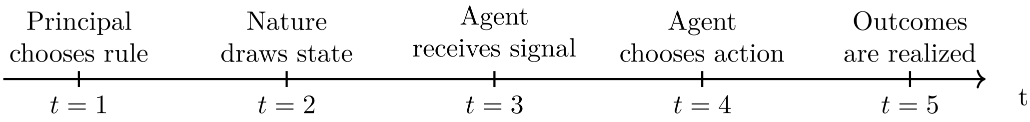


Figure 2.1: Optimal Rule Design - Timeline of the Model

has a positive externality e on welfare. At $t = 5$, the state of the world and the agent's action are revealed and the agent's payoff and total welfare are realized. The timeline is summarized in Figure 2.1.

1. The Agent At $t = 4$, the agent takes a binary action $a \in \{0, 1\}$ at costs $c(a)$, with $c(1) = c$ and $c(0) = 0$. $a = 1$ is a "prosocial" action as it involves a positive externality on welfare. Accordingly, I call $a = 0$ the "selfish" action. The agent's material payoff P depends on the action and the state of the world ω. There are two potential states of the world, a low and a high state. The low state is denoted by $\omega = 0$ and the high state by $\omega = 1$, with $pr(\omega = 1) = \rho$. The agent's material payoff is defined as

$$P = a \cdot \omega - c(a). \tag{2.1}$$

At the time of choosing a, the agent does not know ω. Instead, he receives a private signal s at $t = 3$. Again, the signal is either low ($s = l$), or high ($s = h$). The signal generating process is given by conditional probabilities $pr(s = h|\omega = 1) = \sigma_h$ and $pr(s = h|\omega = 0) = \sigma_l$, with $\sigma_h > 1/2 > \sigma_l$. Using Bayes' rule, the agent updates his beliefs after observing the signal. Denote the resulting posteriors as $\pi_h = pr(\omega = 1|s = h)$ and $\pi_l = pr(\omega = 1|s = l)$. As $\sigma_h > 1/2 > \sigma_l$, it also holds that $\pi_h > \rho > \pi_l$.

1a. The Rule-Following Type The agent is one of two types, $i \in \{F, S\}$ with $pr(i = F) = \mu$. F denotes the "rule-following" type. Besides utility from the material payoff, this type feels an intrinsic disutility when not complying to rules. Let γ_F denote the agent's weight on the concern to follow the rule and $b(a, \tilde{a}(R, s))$ reflect the disutility when breaking a rule. F's utility function is given by

$$u^F = P + \gamma_F \cdot b(a, \tilde{a}(R, s)). \tag{2.2}$$

The disutility from rule-breaking, $b(a, \tilde{a}(R, s))$, depends on the agent's action a as well as action $\tilde{a}(R, s)$ that is dictated by rule R for the agent's private signal s. The functional form of b is given by

$$b(a, \tilde{a}(R, s)) = \begin{cases} 0 & \text{if } a \geq \tilde{a}(R, s) \\ a - \tilde{a} & \text{if } a < \tilde{a}(R, s). \end{cases} \tag{2.3}$$

The agent is neutral to a positive deviation from the rule, i.e. when he chooses the costly action $a = 1$ even though $\tilde{a} = 0$. However, choosing $a = 0$ when the rule dictates $\tilde{a} = 1$ gives a disutility of $-\gamma_F$.

1b. The Selfish Type S denotes the "selfish, image-concerned" type. This type cares about his image and wants to be perceived as a rule-follower. Thus, his utility is increasing in the posterior belief $\hat{\mu}(a, \omega) = pr(i = F|a, \omega)$ of a neutral observer. The observer observes action a and state ω at the end of the game and updates his prior about the agent's type according to Bayes' rule. The observer does not observe the agent's private signal. Let γ_I denote the agent's weight on image concerns. S's utility is given by

$$u^S = P + \gamma_I \cdot \hat{\mu}(a, \omega). \tag{2.4}$$

2. Welfare At $t = 1$, the principal chooses a rule to maximize welfare W. Thereby, she takes into account the positive externality e the agent's action has on welfare. Welfare is not affected by the agent types' intrinsic utility (i.e. the rule-following or image utility) and is given by

$$W = (1 + e) \cdot a \cdot \omega - c(a). \tag{2.5}$$

Moreover, I impose that

$$(1 + e) \cdot \pi_h > (1 + e) \cdot \rho > c > 1 > (1 + e) \cdot \pi_l. \tag{2.6}$$

Implications are as follows. First, $(1 + e) \cdot \rho > c$ implies that from an ex-ante welfare perspective, it is optimal to act prosocially and play $a = 1$. Secondly, $(1 + e) \cdot \pi_h > c > (1 + e) \cdot \pi_l$ implies that from an ex-post welfare perspective, it is optimal to play $a = 1$ if and only if $s = h$, i.e. to take the prosocial action if and only if the signal is high. Thirdly, $c > 1$ implies that the selfish action, $a = 0$, maximizes monetary payoffs regardless of the state or signal. Therefore, if no rule was in place, a conflict of interests would arise between the agent, who then maximizes only his monetary payoff, and the welfare-maximizing principal.

3. Rules To overcome this conflict of interests and to induce the agent to take the welfare-maximizing action, the principal designs a non-binding rule at $t = 1$. There exist no formal sanctions for non-compliance with any rule.

A rule R prescribes an action $\tilde{a}$ that is ought to be taken by the agent after receiving signal s. The principal can choose between the following two rules.[4] The Unconditional Rule R^U dictates the ex-ante efficient behavior, i.e. $\tilde{a} = 1$ regardless of the signal. The Conditional Rule R^C dictates the ex-post efficient behavior, i.e. $\tilde{a} = 1$ if and only if $s = h$. Formally, the rules are given by

$$R^U: \quad \tilde{a} = \quad 1 \;\forall\; s \tag{2.7}$$

$$R^C: \quad \tilde{a} = \begin{cases} 1 & \text{if } s = h \\ 0 & \text{if } s = l. \end{cases} \tag{2.8}$$

2.2.2 Analysis

The analysis focuses on a "wiggle room equilibrium" and shows its existence and uniqueness under certain conditions. In such an equilibrium, the Unconditional Rule induces adherence by both types while under the Conditional Rule, the selfish, image-concerned type does not follow that rule. Definition 2.1 states that intuition formally.

[4]Note that there are generally four rules available. In addition to the two analyzed, there is an Unconditional Rule that prescribes $\tilde{a} = 0 \;\forall\; s$ and a Conditional Rule that prescribes $\tilde{a} = 0(1)$ if $s = h(l)$. Both rules are not included in the analysis because they are clearly inferior to R^U and R^C.

Definition 2.1.
A "wiggle room equilibrium" is a Perfect Bayesian Equilibrium (PBE), using the Intuitive Criterion (Cho and Kreps 1987) as an equilibrium refinement. In this equilibrium the agent behaves as follows:

$$a^F = \quad \tilde{a} \;\; \forall \;\; R \in \{R^U, R^C\} \tag{2.9}$$

$$a^S = \begin{cases} \tilde{a} & \text{if} \quad R = R^U \\ 0 & \text{if} \quad R = R^C, \end{cases} \tag{2.10}$$

and posterior beliefs are updated accordingly.

I proceed by analyzing behavior of the two types separately, starting with the rule-following type. This type is intrinsically motivated to follow any rule since he feels a disutility when not doing so. Therefore, he plays a game of complete information. Lemma 2.1 says that the rule-following type follows any rule and acts prosocially if and only if his rule-following concerns are sufficiently pronounced.

Lemma 2.1.
In any Perfect Bayesian Equilibrium, the rule-following type follows any rule if and only if $\gamma_F \geq c - \pi_l$.

Proof. I first analyze the behavior of the rule-following type under the Unconditional Rule. If $s = h$, rule-following (i.e. $a = 1$) yields an expected utility of $\pi_h - c < 0$. If $s = l$, rule-following gives $\pi_l - c < 0$. In both cases, not following the rule implies that the agent feels a disutility of $-\gamma_F$. As $\pi_l < \pi_h$, (a) the rule-following type follows the Unconditional Rule and acts prosocially after any signal if and only if $\gamma_F \geq c - \pi_l$; (b) if $\gamma_F \in (c - \pi_l, c - \pi_h]$, the agent acts prosocially if and only if $s = h$; (c) if $\gamma_F < c - \pi_h$, the agent never follows the rule and acts selfishly. Under the Conditional Rule, the agent always complies with the rule if $s = l$ since the rule prescribes $a = 0$ and thus there is no conflict between the rule and the payoff-maximizing action. If $s = h$, the same logic applies as under the Unconditional Rule. The agent complies with the rule if and only if $\gamma_F \geq c - \pi_h$. As $\pi_h > \pi_l$, it is implied by $\gamma_F \geq c - \pi_l$ and thus the agent follows both rules if and only if $\gamma_F \geq c - \pi_l$. □

In what follows, I restrict the analysis to the case described by Lemma 2.1.[5] I proceed by analyzing the behavior of the selfish type under the Unconditional Rule, described in Lemma 2.2. It shows the existence of a unique Perfect Bayesian Equilibrium in which the selfish type fully follows the Unconditional Rule. For all proofs throughout this section, it will be useful to define $E_s\hat{\mu}(a) := \pi_s\hat{\mu}(a,1) + (1-\pi_s)\hat{\mu}(a,0)$ as the agent's expectations about the observer's posterior belief when the agent receives signal s and plays action a. Note that, as posterior beliefs are formed in equilibrium, $E_s\hat{\mu}(a)$ is conditional on the equilibrium under consideration.

Lemma 2.2.
Suppose $\gamma_F \geq c - \pi_l$ and $\gamma_I \geq \frac{c-\pi_l}{\mu}$. There exists a unique Perfect Bayesian Equilibrium, using the Intuitive Criterion as equilibrium refinement, in which the selfish type acts prosocially under the Unconditional Rule, regardless of the signal he receives.

Proof. In the unique PBE described in Lemma 2.2, the selfish type always follows the rule. In this case, the selfish type perfectly mimics the rule-following type, and therefore actions contain no information about the agent; hence $\hat{\mu} = \mu$. That gives an expected utility of $\pi_h - c + \gamma_I\mu$ after $s = h$, or $\pi_l - c + \gamma_I\mu$ after $s = l$. Further, note that for the rule-following type $a = 0$ is dominated by $a = 1$ because of Lemma 2.1. The Intuitive Criterion (Cho and Kreps 1987) can be applied: upon observing a deviation to $a = 0$ the observer infers that $\hat{\mu} = 0$. Therefore, only if $\pi_l - c + \gamma_I\mu \geq 0$, i.e. $\gamma_I \geq \frac{c-\pi_l}{\mu}$, the selfish agent follows the Unconditional Rule. This establishes the necessary and sufficient condition. For uniqueness, I first find conditions for all other potential (three pure-strategy and one mixed-strategy) PBE, and then show that these are not consistent with $\gamma_I \geq \frac{c-\pi_l}{\mu}$.

In the second PBE, the selfish type never follows the rule. This gives an utility of 0, while deviating gives $\pi_h - c + \gamma_I$ or $\pi_l - c + \gamma_I$ respectively. Therefore, only if $\gamma_I < c - \pi_h$, the selfish agent never follows the Unconditional Rule. As $c - \pi_h < c - \pi_l$, this is inconsistent with $\gamma_I \geq \frac{c-\pi_l}{\mu} > c - \pi_h$.

In the third PBE, the selfish agent only acts prosocially after $s = h$. This gives a utility of $\pi_h - c + \gamma_I E_h\hat{\mu}(1)$ whereas deviating gives 0 after observing $s = h$. Moreover, in that PBE, the selfish agent plays $a = 0$ after $s = l$, receiving 0. In that case, deviation utility is given by $\pi_l - c + \gamma_I E_l\hat{\mu}(1)$. Therefore, the selfish agent acts prosocially after $s = h$ and selfish after $s = l$ only if $\frac{c-\pi_h}{E_h\hat{\mu}(1)} < \gamma_I < \frac{c-\pi_l}{E_l\hat{\mu}(1)}$. Note that $E_l\hat{\mu}(1) = \pi_l\hat{\mu}(1,1) + (1-\pi_l)\hat{\mu}(1,0) > \mu$ because both posterior beliefs are larger than μ. Thus $\frac{c-\pi_l}{\mu} > \frac{c-\pi_l}{E_l\hat{\mu}(1)}$ and this PBE is inconsistent with $\gamma_I \geq \frac{c-\pi_l}{\mu}$.

[5] In the other two cases, described in the proof of Lemma 2.1, the rule-following type follows the rule either partially or not at all. However they are inconsistent with a wiggle room equilibrium. The equilibrium analysis for these cases can be found in Footnote 6.

The fourth potential PBE states that the selfish agent only acts prosocially after $s = l$. Similar to above, the necessary condition is given by $\frac{c-\pi_l}{E_l\hat{\mu}(1)} < \gamma_I < \frac{c-\pi_h}{E_h\hat{\mu}(1)}$. However, since $\pi_h > \pi_l$, that condition requires that $E_h\hat{\mu}(1) < E_l\hat{\mu}(1)$. Noting $\hat{\mu}(1,1) > \hat{\mu}(1,0)$, $E_s\hat{\mu}(1) = \pi_s\hat{\mu}(1,1) + (1-\pi_s)\hat{\mu}(1,0)$, and $\pi_h > \pi_l$ it becomes apparent that $\frac{c-\pi_l}{E_l\hat{\mu}(1)} > \frac{c-\pi_h}{E_h\hat{\mu}(1)}$. Hence, this case is not a PBE.

Lastly, consider a mixed strategy PBE in which the agent plays $a = 1$ with probability q_h (q_l) $\in (0,1)$ after receiving $s = h$ ($s = l$). In such equilibrium, the agent must be indifferent between playing $a = 1$ and $a = 0$ under each signal. Furthermore, because the rule-following type never acts selfishly (but the selfish type may as $q_l, q_h < 1$), posterior beliefs after $a = 0$ are given by $\hat{\mu}(0,\omega) = 0$. It follows that $E_s\hat{\mu}(0) = 0 \ \forall s$. Therefore, for the agent to be indifferent between $a = 1$ and $a = 0$, it must hold that $\pi_h - c + \gamma_I E_h\hat{\mu}(1) \stackrel{!}{=} \gamma_I E_h\hat{\mu}(0) = 0$ and $\pi_l - c + \gamma_I E_l\hat{\mu}(1) \stackrel{!}{=} \gamma_I E_l\hat{\mu}(0) = 0$ since the rule-following type follows the rule. As $E_s\hat{\mu}(1) = \pi_s\hat{\mu}(1,1) + (1-\pi_s)\hat{\mu}(1,0)$, these two conditions give a system of equations with two unknowns and two equations. Solving gives $\hat{\mu}(1,1) = \frac{c-1}{\gamma_I}$ and $\hat{\mu}(1,0) = \frac{c}{\gamma_I}$. Posterior beliefs are determined in the mixed-strategy equilibrium. Therefore, it must also hold that $\hat{\mu}(1,1) = \frac{\mu}{\mu+(1-\mu)[\sigma_h q_h+(1-\sigma_h)q_l]}$ and $\hat{\mu}(1,0) = \frac{\mu}{\mu+(1-\mu)[\sigma_l q_h+(1-\sigma_l)q_l]}$. Note that, since $q_h, q_l > 0$, both posterior beliefs are larger than μ Hence in the mixed strategy PBE it must jointly hold that $\hat{\mu}(1,1) = \frac{c-1}{\gamma_I} > \mu$ and $\hat{\mu}(1,0) = \frac{c}{\gamma_I} > \mu$, and thus $\frac{c-1}{\mu} > \gamma_I > \frac{c}{\mu}$. The first inequality violates $\gamma_I \geq \frac{c-\pi_l}{\mu}$. Therefore, the pure strategy PBE described in Lemma 2.1 is unique for $\gamma_I \geq \frac{c-\pi_l}{\mu}$. □

Lemma 2.2 states that if image concerns are sufficiently strong even the selfish type will act prosocially and follow the Unconditional Rule. Since the Unconditional Rule prescribes to play $a = 1$ independent of the signal, there is no uncertainty about the agent's rule-following intentions. Hence the two objectives of the selfish, image-concerned type conflict. If image concerns are sufficiently strong, the selfish type rather pools with the rule-following type to maintain a positive image and also follows the rule.

Furthermore, the equilibrium is unique for the range of parameters in which the selfish type acts prosocially under the Unconditional Rule. The underlying intuition is that if the selfish type acts selfishly the posterior beliefs after a prosocial action are higher. That implies that the incentives to act prosocially increase. If image concerns are sufficiently strong, as in Lemma 2.2, any selfish action implies too strong incentives to deviate to a prosocial action and receive image utility. In that case there cannot be any selfish behavior in equilibrium, not even in mixed strategies.

I proceed by analyzing the selfish type's behavior under the Conditional Rule. Lemma 2.3 states that the only existing Perfect Bayesian Equilibrium is one in which the selfish type acts selfishly. However, for that equilibrium to exist image concerns must be sufficiently low. For the analysis, let $g(\mu) := \frac{(1-\mu)[1-\mu(\sigma_l\pi_h+(1-\pi_h)\sigma_h)]}{(1-\mu\sigma_h)(1-\mu\sigma_l)}$, with $g(\mu) \in (0,1)$ if $(\sigma_h < 1 \ \vee \ \sigma_l > 0)$ and $\mu > 0$.

Lemma 2.3.

Suppose $\gamma_F \geq c - \pi_l$ *and* $\gamma_I \leq \frac{c-\pi_h}{g(\mu)}$. *There exists a unique Perfect Bayesian Equilibrium in which the selfish type acts selfishly under the Conditional Rule. If* $\gamma_I > \frac{c-\pi_h}{g(\mu)}$, *there does not exist a Perfect Bayesian Equilibrium under the Conditional Rule, neither in pure nor in mixed strategies.*

Proof. Before proving the existence of the PBE in Lemma 2.3, I prove that no other Perfect Bayesian Equilibrium exists.

First, consider a potential PBE in which the selfish type acts prosocially. In that case, only the prosocial type may act selfishly, thus $\hat{\mu}(a=0) = 1$. Hence a deviation of the selfish type to $a = 0$ increases both the monetary payoff ($a = 1$ gives $\pi_s - c < 0$) *and* the type's image. Thus that case cannot constitute an equilibrium.

Secondly, consider a potential PBE in which the selfish type acts prosocially only if $s = h$. Note that in this equilibrium, the selfish type perfectly mimics the rule-following type. Therefore, actions contain no information regarding the agent's type and any posterior belief is equal to the prior. After $s = h$, the selfish type plays $a = 1$ which results in an expected utility of $\pi_h - c + \gamma_I \mu$. Deviation gives $\gamma_I \mu$. As $\pi_h - c < 0$, the selfish type deviates.

Thirdly, consider a potential PBE in which the selfish type acts prosocially only if $s = l$. After $s = h$, he plays $a = 0$ which results in an expected utility of $\gamma_I E_h \hat{\mu}(0)$. Deviating gives $\pi_h - c + \gamma_I E_h \hat{\mu}(1)$. After $s = l$, he plays $a = 1$ which results in an expected utility of $\pi_l - c + \gamma_I E_l \hat{\mu}(1)$. Deviating gives $\gamma_I E_l \hat{\mu}(0)$. The resulting two conditions can be summarized as $\frac{c-\pi_l}{E_l\hat{\mu}(1)-E_l\hat{\mu}(0)} < \gamma_I < \frac{c-\pi_l}{E_h\hat{\mu}(1)-E_h\hat{\mu}(0)}$. However, note that $\hat{\mu}(1,1) > \hat{\mu}(1,0)$ (in a high state, a high action is more likely to be from a rule-following type) and $\hat{\mu}(0,0) > \hat{\mu}(0,1)$ (in a low state, a low action is more likely to be from a rule-following type). As $\pi_h > \pi_l$ and $E_s\hat{\mu}(a) = \pi_s\hat{\mu}(a,1) + (1-\pi_s)\hat{\mu}(a,0)$, it follows that $E_l\hat{\mu}(1) - E_l\hat{\mu}(0) < E_h\hat{\mu}(1) - E_h\hat{\mu}(0)$ and thus it can never hold that $\frac{c-\pi_l}{E_l\hat{\mu}(1)-E_l\hat{\mu}(0)} < \frac{c-\pi_l}{E_h\hat{\mu}(1)-E_h\hat{\mu}(0)}$.

Fourthly, consider a potential mixed strategy PBE in which the agent plays $a = 1$ with probability q_h (q_l) $\in (0,1)$ after receiving $s = h$ ($s = l$). In such equilibrium, the agent must be indifferent between playing $a = 1$ and $a = 0$ under each signal. Therefore, it must hold that $\pi_h - c + \gamma_I E_h\hat{\mu}(1) = \gamma_I E_h\hat{\mu}(0)$ and $\pi_l - c + \gamma_I E_l\hat{\mu}(1) = \gamma_I E_l\hat{\mu}(0)$. Putting the two conditions together, it must hold that $\pi_h + \gamma_I[E_h\hat{\mu}(1) - E_h\hat{\mu}(0)] = \pi_l + \gamma_I [E_l\hat{\mu}(1) - E_l\hat{\mu}(0)]$. $\pi_h > \pi_l$ implies that $E_h\hat{\mu}(1) - E_h\hat{\mu}(0) < E_l\hat{\mu}(1) - E_l\hat{\mu}(0)$ is a necessary condition for a mixed strategy equilibrium. I can re-write that condition as $\pi_h\left[(\hat{\mu}(1,1)-\hat{\mu}(1,0))-(\hat{\mu}(0,1)-\hat{\mu}(0,0))\right] < \pi_l\left[(\hat{\mu}(1,1)-\hat{\mu}(1,0))-(\hat{\mu}(0,1)-\hat{\mu}(0,0))\right]$. Therefore, since $\pi_h > \pi_l$, it must hold that $\left[(\hat{\mu}(1,1)-\hat{\mu}(1,0))-(\hat{\mu}(0,1)-\hat{\mu}(0,0))\right] < 0$. However, as the rule-following type follows the Conditional Rule perfectly, in any mixed strategy equilibrium it holds that $\hat{\mu}(1,1) > \hat{\mu}(0,1)$ and/or $\hat{\mu}(0,0) > \hat{\mu}(1,0)$, i.e. the posterior belief is higher after observing a rule-following action. Therefore, $\left[(\hat{\mu}(1,1)-\hat{\mu}(1,0))-(\hat{\mu}(0,1)-\hat{\mu}(0,0))\right]$ is positive and the necessary condition for a mixed strategy PBE is violated.

The last remaining Perfect Bayesian Equilibrium is the one in which the selfish agent acts selfishly. That gives an expected utility of $\gamma_I E_h\hat{\mu}(0)$ and $\gamma_I E_l\hat{\mu}(0)$ after observing $s=h$ and $s=l$ respectively. Deviating to $a=1$ gives $\pi_h - c + \gamma_I$ and $\pi_l - c + \gamma_I$ as $\hat{\mu}(a=1)=1$. Therefore, within that equilibrium it must hold that $\gamma_I \leq \frac{c-\pi_h}{1-E_h\hat{\mu}(0)}$ and $\gamma_I \leq \frac{c-\pi_l}{1-E_l\hat{\mu}(0)}$. Note that $\hat{\mu}(0,0) > \hat{\mu}(1,0)$, i.e. a selfish action is more likely to come from a rule-following type if the state is low. As $\pi_h > \pi_l$ it follows that (a) $E_h\hat{\mu}(0) < E_l\hat{\mu}(0)$ and subsequently (b) $\frac{c-\pi_h}{1-E_h\hat{\mu}(0)} < \frac{c-\pi_l}{1-E_l\hat{\mu}(0)}$. Therefore, in the PBE it is necessary and sufficient that $\gamma_I \leq \frac{c-\pi_l}{1-E_l\hat{\mu}(0)}$. Lastly, note that $1-E_h\hat{\mu}(0) = 1-(\pi_h\cdot\frac{\mu(1-\sigma_h)}{1-\mu\sigma_h}+(1-\pi_h)\cdot\frac{\mu(1-\sigma_l)}{1-\mu\sigma_l})$ and re-formulation gives $1-E_h\hat{\mu}(0)=g(\mu)$. That yields the necessary and sufficient condition for the unique PBE described in Lemma 2.3, i.e. $\gamma_I \leq \frac{c-\pi_h}{g(\mu)}$. □

The main intuition behind Lemma 2.3 is straightforward. If image concerns are sufficiently weak, the selfish type acts selfishly under the Conditional Rule. If image concerns were sufficiently strong, the selfish type would have an incentive to act prosocially, deviate and mimic the rule-following type. Moreover the equilibrium in which the selfish type acts fully selfish is unique. In any equilibrium in which the selfish type would act prosocially (or attach a positive probability to such action), the ambiguity that arises with the Conditional Rule creates an incentive to deviate and act selfishly.[6]

2.2.3 A Wiggle Room Equilibrium

As the behavior described in Lemmas 2.1, 2.2 and 2.3 is consistent with the wiggle room equilibrium in Definition 2.1, Proposition 2.1 follows naturally.

[6]Note that the analysis has only considered the selfish type's behavior conditional on $\gamma_F \geq c-\pi_l$. Clearly, the other two cases, i.e. when the rule-following type does not perfectly follow both rules, are inconsistent with a wiggle room equilibrium. Yet for completeness, I state the equilibria for these cases here. If $\gamma_F < c-\pi_h$, there is a unique equilibrium in which no type follows any rule and both types act selfishly. For the intermediate case, $\gamma_F \in [c-\pi_h, c-\pi_l)$, there is a unique equilibrium in which the selfish type acts selfishly if and only if γ_I is sufficiently small. If not, there exists no equilibrium, neither in pure nor in mixed strategies. The proof is similar to the proof of selfish behavior under the Conditional Rule for Lemma 2.3 and is thus omitted.

Proposition 2.1.

Suppose $\gamma_F \geq c - \pi_l$ *and* $\frac{c-\pi_h}{g(\mu)} \geq \gamma_I \geq \frac{c-\pi_l}{\mu}$. *The wiggle room equilibrium exists and is unique.*

Proof. The conditions follow from Lemmas 2.1, 2.2 and 2.3. It is left to prove that there exist μ such that $\frac{c-\pi_h}{g(\mu)} > \frac{c-\pi_l}{\mu}$, i.e. $g(\mu) < \mu$ and thus $\frac{c-\pi_h}{g(\mu)} \geq \gamma_I \geq \frac{c-\pi_l}{\mu}$ is possible. That proof is relegated to the proof of Condition 4 of Corollary 2.1.

Proposition 2.1 summarizes the necessary and sufficient conditions for the existence of a wiggle room equilibrium. It reflects that, within the wiggle room equilibrium, no type can have an incentive to deviate. As stated in Lemma 2.1, the rule-following type adheres to both rules if and only if his intrinsic motivation for rule-following is sufficiently high. Lemmas 2.2 and 2.3 show that the selfish type follows the Unconditional Rule if his image concerns are sufficiently strong, but acts selfishly under the Conditional Rule if his image concerns are sufficiently weak. Therefore image concerns must be intermediate. Furthermore, from Proposition 2.1 follow further restrictions on the signal structure and type composition. Corollary 2.1 summarizes four necessary conditions for the wiggle room equilibrium to exist.

Corollary 2.1.

A wiggle room equilibrium exists only if

1. $\gamma_F \geq c - \pi_l$: *rule-following concerns are sufficiently pronounced, and*

2. $\gamma_I \in [\frac{c-\pi_l}{\mu}, \frac{c-\pi_h}{g(\mu)}]$: *image concerns are intermediate, and*

3. $\sigma_h < 1 \;\vee\; \sigma_l > 0$: *signals are noisy, and*

4. $\mu > \underline{\mu}$: *the fraction of rule-following agents is sufficiently large.*

and $\underline{\mu} = \max\left\{\frac{c-\pi_l}{\gamma_I}, \mu_1\right\}$ *and* μ_1 *is implicitly given by* $g(\mu_1) \overset{!}{=} \mu_1$.

Proof. Conditions 1 and 2 are stated in Proposition 2.1 already. For Condition 3, note that the wiggle room equilibrium cannot exist if $g(\mu) = 1$ as in that case $\frac{c-\pi_l}{\mu} > \frac{c-\pi_h}{g(\mu)}$, a violation of Condition 2. Recall that $g(\mu) = \frac{(1-\mu)[1-\mu(\sigma_l\pi_h+(1-\pi_h)\sigma_h)]}{(1-\mu\sigma_h)(1-\mu\sigma_l)}$. Suppose $\sigma_h = 1$ and $\sigma_l = 0$. Then $g(\mu) = 1 - (1 - \pi_h)\mu$. However, $\sigma_l = 0$ also implies that $\pi_h = 1$ and thus $g(\mu) = 1$. Therefore, if $\sigma_h = 1$ and $\sigma_l = 0$, the wiggle room equilibrium does not exist. However, if $\sigma_h = 1$ and $\sigma_l > 0$,

$\pi_h < 1$ and $g(\mu) = \frac{1-\mu-\mu\pi(1-\sigma_l)}{(1-\mu\sigma_l)} \in (0,1)$. If $\sigma_h < 1$ and $\sigma_l = 0$, $\pi_h = 1$ and $g(\mu) = \frac{(1-\mu)[1-\mu(1-\pi_h)\sigma_h]}{1-\mu\sigma_h} \in (0,1)$.

The first part of Condition 4, $\mu \geq \frac{c-\pi_l}{\gamma_I}$ emerges from $\gamma_I \geq \frac{c-\pi_l}{\mu}$. For the second part of Condition 4, i.e. $\mu \geq \mu_1$, note that $\frac{c-\pi_l}{\mu} \leq \gamma_I \leq \frac{c-\pi_h}{g(\mu)}$ implies that it must hold that $g(\mu) < \mu$ as $\pi_h > \pi_l$. Now note that if Condition 3 holds, $g(\mu) \in (0,1)$ and it is continuous in μ. Secondly, also $\mu \in [0,1]$. Therefore, Brouwer's fixed theorem states that μ_1 exists with $g(\mu_1) \stackrel{!}{=} \mu_1$. Lastly, to show that μ_1 is a lower bound to μ, note that $g(\mu)$ is decreasing in μ, with $g(0) = 1$. Hence, for $\mu > g(\mu)$, it must hold that $\mu > \mu_1$. □

The first two conditions of Corollary 2.1 have been discussed already. The third condition describes that a wiggle room equilibrium, quite naturally, builds on signals' noisiness. Formally, the wiggle room equilibrium requires that $g(\mu) < 1$, which implies that the expected posterior beliefs in equilibrium, $E_h\hat{\mu}(0)$, are positive. Intuitively, signals must be noisy so that the selfish type can maintain a positive image to some degree. That nosiness is captured by $g(\mu) < 1$. However, if both signals are precise it holds that $g(\mu) = 1$, and actions perfectly reveal the agent's type; hence there is no possibility for a positive image if acting selfishly. That gives an incentive for the selfish type to deviate and act prosocial to fully recover his image. Therefore, with precise signals a wiggle room equilibrium cannot exist.

However, it is not required that both signals are noisy. If $\sigma_h < 1$ and $\sigma_l = 0$ the agent may receive a low signal even though the state is high. In this case, despite a high state and a selfish action, the agent can have still followed the rule as he may have received the misleading, low signal. Thus, the selfish type can use this possibility to act selfishly under the Conditional Rule. If $\sigma_h = 1$ and $\sigma_l > 0$ the agent may receive a high signal even though the state is low. Here, the selfish type "gambles" and hopes that the signal is actually misleading. If the state is low despite a high signal, his selfish action would not be socially punished (as the rule-follower also plays $a = 0$ when the signal and state are low). Therefore, the existence of some noise is necessary for the emergence of a moral wiggle room. However, the selfish type's decision to use it and act selfishly depends on the relative size of his image concerns.

The fourth condition of Corollary 2.1 describes that there must be sufficiently many rule-followers. Otherwise, any posterior belief that the agent is a rule-following type is too low and the selfish type has no incentive to act prosocially.

Hence, the existence of a wiggle room equilibrium is quite sensible to exact parameter constellations.[7] This is one of the major complications for the experimental setup, described in Section 2.3. I conclude the theory section by examining the welfare implications under both rules in a wiggle room equilibrium.

Welfare

To analyze welfare consequences, denote W^U as the expected welfare under the Unconditional Rule R^U and W^C the expected welfare under the Conditional Rule R^C. In the wiggle room equilibrium they are given by

$$W^U = \rho \cdot (1+e) - c \tag{2.11}$$

$$W^C = \mu \cdot [\rho\sigma_h \cdot (1+e-c) - (1-\rho)\sigma_l c]. \tag{2.12}$$

Under the Unconditional Rule, any agent plays $a = 1$ at costs c. Only if $\omega = 1$, i.e. with a probability of ρ, that action translates into welfare benefits of $(1+e)$.

Under the Conditional Rule the selfish type always takes the selfish action, inducing a welfare of zero. The agent is a rule-following type with probability μ. In that case, he plays $a = 1$ at costs c if and only if the signal is high. If the state turns out to be high as well (with probability $\rho\sigma_h$), there is a welfare gain of $(1+e)$ and costs of c. If the state is low despite a high signal (with probability $(1-\rho)\sigma_l$), the prosocial action does not pay off. Comparing both welfares under the two rules gives Proposition 2.2.

Proposition 2.2.
In the wiggle room equilibrium, the Unconditional Rule induces a higher welfare only if

$$\frac{1+e}{c} - 1 > \frac{1-\rho}{\rho}\frac{1-\mu\sigma_l}{1-\mu\sigma_h}. \tag{2.13}$$

[7] Interestingly, in any other existing equilibrium there is even *less* rule-following under the Unconditional Rule, see Footnote 6 and the proof for Lemma 2.2.

Proof. The proof involves only some mathematical reformulation of $W^U > W^C$ and is thus omitted. □

First, it is apparent that the Unconditional Rule becomes more likely if the prosocial action's externality increase and costs decrease. In that case, the welfare benefits from the prosocial action in a high stage increase; thus the relative inefficiency of the Unconditional Rule (i.e. that it prescribes the inefficient action for $s = l$) decreases.

Secondly, to analyze the effect of the wiggle room on the relative welfare benefits of the Unconditional Rule, I focus on the comparative statics of μ on $W^U - W^C$ under these two conditions. Suppose (a) the conditions in Proposition 2.1 hold, i.e. a wiggle room equilibrium exists, and (b) that if all agents followed the rule the Conditional Rule would induce higher welfare, i.e. $\frac{1+e}{c} - 1 < \frac{1-\rho}{\rho}\frac{1-\sigma_l}{1-\sigma_h}$. This is the case if $(1 - \sigma_l)$ and σ_h are sufficiently high, i.e. the signals' noisiness is sufficiently low.

Proposition 2.2 implies an upper bound $\overline{\mu}$ on the share of rule-following individuals for the Unconditional Rule to be superior. For the intuition, suppose there are only rule-following individuals, i.e. $\mu = 1$. In that case, the Conditional Rule is superior as it prescribes the efficient action for each signal. However, in the wiggle room equilibrium selfish types do not follow the Conditional Rule. That relative disadvantage increases when μ decreases; less rule-following types implies more selfish behavior only under the Conditional Rule. Because both types follow the Unconditional Rule, it becomes superior if the Conditional Rule cannot induce sufficiently many individuals to act prosocially, i.e. if $\mu < \overline{\mu}$. This finding is stated in Corollary 2.2.[8]

Corollary 2.2.
Suppose $c - \pi_l \leq \gamma_F$, $\frac{c-\pi_l}{\mu} \leq \gamma_I \leq \frac{c-\pi_h}{g(\mu)}$ *and* $\frac{1+e}{c} - 1 < \frac{1-\rho}{\rho}\frac{1-\sigma_l}{1-\sigma_h}$. *The Unconditional Rule induces higher welfare if and only if* $\mu \in [\underline{\mu}, \overline{\mu})$.

Proof. The proof involves only to solve Equation (2.13) for μ and noting that Corollary 2.1 states a lower bound on μ and is thus omitted. □

[8]Here, $\underline{\mu}$ is defined by Corollary 2.1 and $\overline{\mu} = \frac{\rho(1+e)-c}{\rho\sigma_h(1+e)-(\rho\sigma_h+(1-\rho)\sigma_l)c}$.

2.3 Experiment

An experiment was conducted to test two main questions. First, do individuals follow rules in an uncertain environment? The theory clarifies that this is a necessary condition for the existence of a wiggle room equilibrium. The second question builds on that condition. Does a wiggle room equilibrium emerge in the experiment, i.e. do individuals act more selfishly under the Conditional Rule?

2.3.1 Design

The experiment uses a variant of a three-person public goods game (PGG). It consists of four parts. In Part I, subjects play the PGG three times for different multiplication factors. In the main part of the experiment, uncertainty as well as rules and (self-)image concerns are introduced to the PGG. It consists of a one-shot (Part II) and repeated version (Part III) of the PGG. In Part IV, a post-experimental survey is conducted. Details about the survey as well as a summary of the experimental structure (Figure 2.3) are to be found at the end of this section.

Part I

In Part I, subjects are randomly matched into groups of three. In a group, each subject receives an endowment of 20 points and decides over a contribution $x \in \{0, 10, 20\}$ to the group account. Every point that is not contributed to the group account is allocated to a private account. A subject earns one point for each point allocated to his/her private account. A contribution to the group account of x points translates into an earning of y points for each of the three subjects in the group. The exact amount y depends on the level of contribution and the state of the world ω that applies to one group. There are three potential states, called State 1, State 2 and State 3.[9] The relationship between contribution x and earnings y for each state can be found in Table 2.1. For example, for one contribution of $x = 20$ points, each subject in the group receives $y = 12$ (9, 5) points if the state is $\omega =$ State 1 (2, 3). Given a

[9] In the experiment, states were called "situations". All instructions as well as exemplary screenshots can be found in Appendix B.2.

ω	State 1	State 2	State 3
$x = 0$	$y = 0$	$y = 0$	$\underline{y = 0}$
$x = 10$	$y = 5$	$\underline{y = 8}$	$y = 2$
$x = 20$	$\underline{y = 12}$	$y = 9$	$y = 5$

Table 2.1: Earnings Scheme for the Group Account

state ω, full earnings for individual i are given by

$$\pi_i = 20 - x_i + \sum_{j=1}^{3} y_j(x_j, \omega). \tag{2.14}$$

For all states, the payoff-maximizing contribution is $x^S = 0$. But states differ in their efficient, welfare-maximizing contribution x^*. The respective efficient contributions are underlined in Table 2.1. In State 1, the efficient contribution is $x^* = 20$ points, in State 2 it is $x^* = 10$ points and in State 3, $x^* = 0$.

In Part I of the experiment, subjects play the PGG for each state separately. After explaining the general public goods game, the relationship between a contribution and the subjects' earnings for State 1 is introduced first. Control questions follow. Then, subjects simultaneously make their contribution decision for State 1. Thereafter, beliefs about the other group members' contributions are elicited. That procedure is repeated for State 2 and State 3.[10] There is no feedback on others' contributions within or after Part I. Only at the end of the experiment, one of the three states is randomly drawn by the computer and earnings are distributed according to subjects' contributions and the drawn state.

In Part I, subjects are therefore introduced to the general trade-off of a PGG in an incentivized manner. Due to the complex environment in the main part of the experiment, it is important that subjects understand the PGG and its implications in each state.

Main Part - Parts II and III

The main part of the experiment introduces uncertainty as well as rules. It consists of a one-shot (Part II) and repeated (Part III) version of the following three-person PGG; I explain the one-shot version first.

[10] In half of the sessions, the order of states is reversed.

After Part I, subjects are randomly re-matched into groups of three and randomly assigned one of two roles. There is one group that consists of "rulers" who, in Part II, choose non-binding rules for other groups. The remaining groups consist of "contributors".

Contributors play a PGG similar to Part I. Endowments, states and the earnings scheme from Table 2.1 remain the same. The overall structure of the one-shot PGG in Part II is described in Figure 2.2. At $t = 1$, the computer privately draws a state that determines the relationship between contributions and earnings for each group. Based on that draw, the computer further draws private independent signals for each subject of a group. At $t = 2$, rulers choose a rule for each group. At $t = 3$, contributors are informed about the rule in their group. Thereafter, contributions are elicited via the strategy method, i.e. for each signal, at $t = 4$. This part's payoffs and contribution are revealed to the subjects only at the end of the experiment.

Uncertainty and Signals. At $t = 1$, the computer privately draws one of the three states with equal probability. Subjects do not learn the relevant state until the end of the experiment. But they do receive a noisy, private signal about the computer draw. There are three potential signals, Signal I, Signal II and Signal III. After the state is drawn, the computer draws a signal for each subject in a group independently. The signal matches the state with probability 5/6.[11] Each of the other two, "misleading" signals is drawn with a probability of 1/12. Subjects are instructed about the signal structure as well as the resulting posteriors. The posteriors are depicted using a pie chart to ease understanding. Subjects further learn that signals are private and no other subject can observe other subjects' signals.

To compare behavior under an Unconditional Rule and a Conditional Rule, all three subjects within a group receive the same non-binding rule that guides behavior. The rule is given to the subjects before their contribution decision. The experiment's treatment variation is therefore on the group level and consists of the two different rules.

[11] That is, Signal I is the matching signal for State 1 etc. The conditional probabilities are chosen such that the efficient contribution remains the same for each state and the according signal.

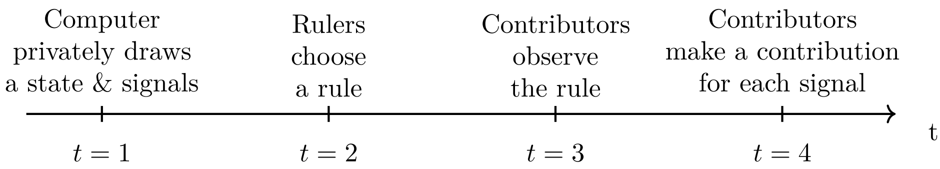


Figure 2.2: Timeline in Part II, the One-Shot PGG.

Rule Choice and Treatment Variation. Before the subjects' contribution decisions, all subjects in a group receive a "group rule". The group rule is chosen by one of the rulers who do not play the PGG, neither in Part II nor Part III. The contributors are instructed that the ruler "has an incentive to choose a group rule that induces a group payoff that is as high as possible." This is "because his own earnings from this part of the experiment depend inter alia on this group payoff." That procedure is ought to make the Unconditional Rule "plausible" and induce sufficient rule-following behavior, even when it prescribes inefficient behavior.

Each ruler is matched with two groups and chooses one group rule for each group. There are three rules, displayed in Table 2.2. Rule A is the Unconditional Rule. It prescribes a contribution of 10 points regardless of the signal. This is the ex-ante efficient contribution. Rule C is the Conditional Rule. It dictates the efficient contribution for each signal. Rule B is, in expected terms, inferior to both rules. It dictates the selfish contribution of zero points for all signals. For the first group the ruler is matched with, he chooses between the Unconditional Rule A and Rule B. For the second group, he chooses between the Conditional Rule C and Rule B. In additional instructions, only available to the rulers, it is explicitly stated that Rule A and Rule C are superior to Rule B as they generate higher expected group if subjects follow the rule. As rulers receive the average payoff of one of the groups randomly, they are incentivized to choose the Unconditional Rule and the Conditional Rule. All subjects were informed that rules are non-binding, i.e. there exist no formal sanctions for non-compliance.

After rulers have chosen a rule for each group, the chosen group rule is depicted to each subject in a group on the computer screen. Subjects know only about the chosen rule; they are informed neither about the existence of other rules nor about the procedure of the ruler's decision.

When receiving ..., contribute:	Unconditional Rule A	Rule B	Conditional Rule C
Signal I	10 points	0 points	20 points
Signal II	10 points	0 points	10 points
Signal III	10 points	0 points	0 points
	Choice for Group 1 (Rule A, Rule B)		
		Choice for Group 2 (Rule B, Rule C)	

Table 2.2: Rules in the Experiment

In conclusion, that setup allows for a treatment variation on the group level. Groups receiving Rule A are in the "Unconditional" treatment (UNCOND). Groups receiving Rule C are in the "Conditional" treatment (COND). Throughout the whole experiment, the only difference between the two treatments is the group rule.

One-shot PGG (Part II). After the general instructions for Part II are read aloud, subjects answer unincentivized control questions. They are assigned to their respective roles as contributors and rulers. After the computer has privately drawn states for the contributor groups and signal for their members, the ruler chooses a group rule (without knowing the draws). The selected group rule is then displayed to all subjects in a group. Thereafter, the subjects' contributions are elicited via the strategy method: subjects make a contribution decision for each signal without knowing the signal that was drawn, first for Signal I, then II and III. Moreover, beliefs about other group members' contributions are elicited after the respective contribution decision. If beliefs are correct, subjects receive additional 5 points. At the end of Part II, subjects are informed about which signal was drawn to determine their contribution. Only at the end of the experiment, subjects learn the state in Part II and the contributions of the other group members (but not the others' private signals).

Moreover, rulers state beliefs about the expected groups' average payoffs. Thereby I measure whether rulers anticipate rule-based changes in the group payoffs and thus welfare differences between the rules. Rulers receive additional 5 points if their belief is correct within a 1-point-range.

Repeated PGG (Part III). The repeated version of the PGG consists of ten rounds. Groups are stable between Part II and III as well as across rounds in Part III. Also the group rule remains the same. At the beginning of each round, subjects receive feedback: the states as well the other group members' contributions in all previous rounds are displayed. Furthermore, each subject (privately) observes his/her signals, contributions and payoffs in all previous rounds. In a current round, the computer first draws a state for the group for that round and then draws independent signals for each subject. Then, subjects decide over their contributions to the group account. The group rule as well as the feedback and the current signal are depicted on the decision screen.

Groups and subjects are matched across treatments (i.e. group rules) on the sequence of states, and the sequence of signals respectively. This allows for a clean comparison of rule effects in the repeated PGG game. For each group, one of the ten rounds was randomly chosen to determine subjects' earnings in Part III.

Rulers' task in Part III. While contributors play the repeated version of the PGG, rulers face a different task. They undertake a belief elicitation task to elicit prevalent social norms induced by the different rules. Thereby, I use simple coordination games based on Krupka and Weber (2013): rulers state their beliefs over another ruler's response on the social appropriateness of contributions. They do so for each of the three different states under both the Unconditional Rule and and the Conditional Rule. This gives six belief measures (3 states x 2 rules). Thereby, they rate hypothetical contributions on a scale of "Socially very inappropriate, Socially inappropriate, Socially appropriate, Socially very appropriate". Rulers are incentivized to match the answer of a randomly chosen other ruler in their group. One of the six answers is randomly chosen for the payoff. If the answer is equal to the matched ruler's answer, the ruler receives 20 points. As described by Krupka and Weber (2013) this method allows to identify prevalent social norms in the situation at hand.

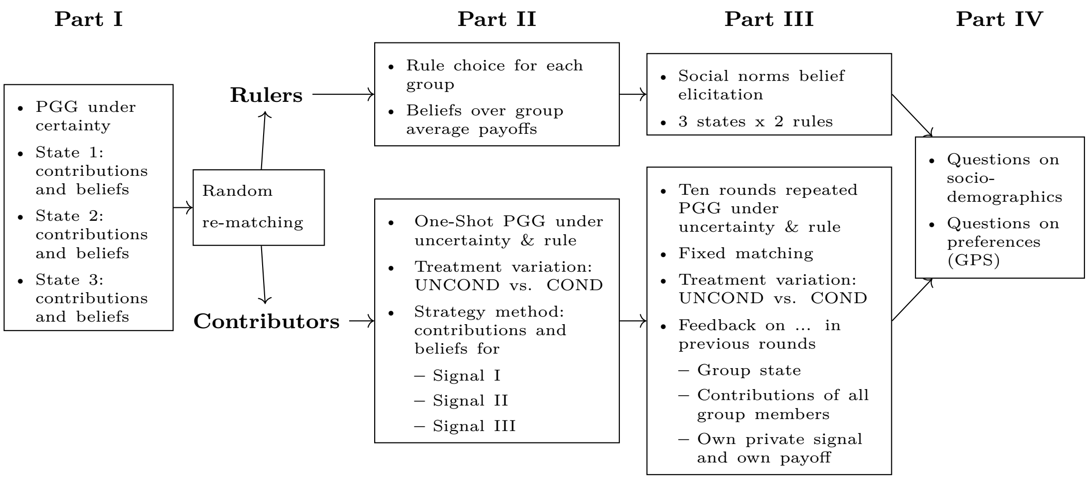


Figure 2.3: Overall Structure of the Experiment

Part IV

At the end of the experiment, all participants complete an (unincentivized) survey containing socio-demographic questions as well as questions on trust, risk, negative reciprocity, altruism and math skills from the Global Preference Survey (Falk et al. 2018). The complete structure of the experiment is presented in Figure 2.3.

2.3.2 Hypotheses

The first hypothesis regards the rulers' rule choice. Treatment variation only works if rulers choose the Unconditional Rule and the Conditional Rule, respectively, over the inferior Rule B.

Hypothesis 2.1.
Rulers prefer the Unconditional Rule and the Conditional Rule over Rule B.

The second hypothesis investigates subjects' rule-following behavior. The summary in Table 2.3 highlights that the Unconditional Rule and the Conditional Rule differ in the prescribed contributions for Signal I and Signal III. Thus, if subjects tend to follow rules contributions in COND will be higher when the signal is Signal I and lower when the signal is Signal III.

	Signal I	Signal II	Signal III
Selfish contribution x^S	0 points	0 points	0 points
Efficient contribution x^*	20 points	10 points	0 points
UNCOND: Unconditional Rule	10 points	10 points	10 points
COND: Conditional Rule	20 points	10 points	0 points

Table 2.3: Contributions and Rules

Hypothesis 2.2.
Subjects tend to follow the rules, i.e. contributions are higher (lower) under the Conditional Rule for Signal I (Signal III).

Given that a substantial fraction of subjects follows the rules, the possibility for a wiggle room exists. In Part III, the social appropriateness of contributions under each rule is elicited from rulers. In terms of the model, social appropriateness is an approximation for social image. The more socially appropriate a contribution is, the higher the contributor's social image should be. In that sense, rulers in Part III take the role of a neutral observer. If they update consistently, rulers should infer that a contribution of zero points is socially *less inappropriate* under the Conditional Rule. This is because under the Unconditional Rule, a contribution of zero points is unambiguously rule-breaking. Under the Conditional Rule, a contribution of zero points can still be rule-following (if the subject has received a misleading Signal III). I focus on the social appropriateness of actions for State 2. In terms of the model, State 2 and Signal II correspond to $\omega = 1$ and $s = h$ as both rules prescribe the same prosocial action in that case. Under the Unconditional Rule, the model predicts $\hat{\mu}(\omega = 1, a = 0) = 0$. Under the Conditional Rule, the model predicts $\hat{\mu}(\omega = 1, a = 0) > 0$. Hypothesis 2.3 applies that intuition to the current experiment.

Hypothesis 2.3.
For State 2, a contribution of zero points is socially less inappropriate under the Conditional Rule.

Given that all three previous hypotheses hold, there is room for wiggle room behavior. If the wiggle room equilibrium exists, subjects will contribute less under the Conditional Rule. As Table 2.3 shows, both rules dictate the same contribution of 10 points when receiving Signal II. Therefore, Signal II allows for a clean comparison of the rules. There, the only difference between the two rules lies in the potential to "wiggle" under the Conditional Rule. Hypothesis 2.4 states that contributions are lower under the Conditional Rule when subjects receive Signal II.

Hypothesis 2.4.
Subjects use the possibility to act selfishly under the Conditional Rule, i.e. contributions are lower under the Conditional Rule for Signal II.

2.3.3 Procedures

The experiment took place at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA), with a total of 10 sessions and 207 participants. A pilot session was conducted to test the subjects' understanding of the experiment.[12] In the experiment, there were 21 participants per session, except for one session with 18 participants. Participants were recruited via ORSEE (Greiner 2015) and the experiment was conducted via oTree (Chen et al. 2016). One point was converted into 0.125 EUR at the end of the experiment. The subjects' earnings from each of the three experimental parts were summed up to determine their total earnings. These ranged between 11 and 21.50 EUR (12.86 and 25.14 US-$), with an average of 15.71 EUR (18.37 US-$). On average, a session lasted around 61 minutes. The experiment was pre-registered.[13]

2.3.4 Main Results

Rule Choice

In the experiment, 30 rulers made 59 rule choices.[14] In 57 of them, the Unconditional or Conditional Rule was chosen, i.e. a percentage of 96.6%.

[12] Due to administrative problems, one additional session had to be aborted after Part I. The data from that session was not used for any analysis.

[13] http://aspredicted.org/blind.php?x=uh7pu3

[14] In one session, there were only 18 participants. Therefore, one ruler had only one matched contributor group and made only one rule choice. It was randomly determined whether the Unconditional Rule or Conditional Rule was the alternative to Rule B.

Result 2.1.
Rulers choose the superior rule.

I exclude all groups that play under Rule B from the analysis. In total 28 groups (84 subjects) act under the Unconditional Rule and 29 groups (87 subjects) under the Conditional Rule.

Rule-Following Behavior

Hypothesis 2.2 states that the Conditional Rule induces higher contributions for Signal I and lower contributions for Signal III. Figure 2.4 shows the mean contributions in Part II and III for each signal under the different rules. When the Conditional Rule dictates a higher contribution for Signal I, subjects contribute significantly more (Part II: p=0.0015. Part III: p=0.0000).[15] Similarly, for Signal III when the Conditional Rule dictates a lower contribution, subjects contribute significantly less (p=0.0000 in both parts). However, subjects follow the rules imperfectly and tend to behave more selfishly than the rule prescribes. Contributions are lower than the rule dictates in all but one case.[16] Moreover, under the Unconditional Rule more subjects follow the rule for Signal I than for Signal III as in the former case rule-following is more in line with selfish behavior for Signal I than for Signal III.[17] In conclusion, Hypothesis 2.2 is confirmed.

Result 2.2.
Contributors tend to follow the rules.

Before examining subjects' wiggle room behavior, I first investigate the existence of a potential wiggle room, based on the social appropriateness of contributions. Figure 2.5 displays the means of the social appropriateness measures elicited from rulers in Part III. Following Krupka and Weber (2013),

[15] Unless noted otherwise, a two-sided Wilcoxon-Mann-Whitney test is used to test for significance. For the analysis in Part III, I further exclude two additional groups that were matched with the two groups under Rule B.

[16] Differences to the rules' prescribed contributions are highly significant at the 1%-level for each signal and in both parts (using a two-sided Student's t-test). Contributions are slightly higher than zero for Signal III under the Conditional Rule.

[17] The proportions are: 45 out of 84 subjects contribute 10 points for Signal I and 32 subjects for Signal III (p=0.0000, Fisher's exact test).

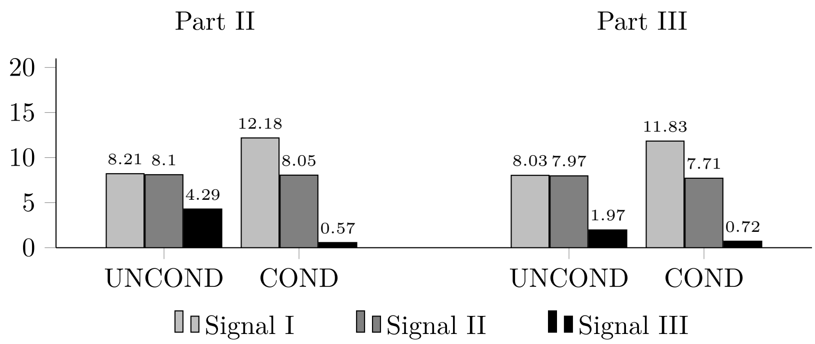


Figure 2.4: Mean Contributions for Each Signal in Part II and Part III

the Likert scale was transformed into values of $(-1, -\frac{1}{3}, \frac{1}{3}, 1)$. Thus a higher mean is associated with a higher perceived social appropriateness.

First, social appropriateness of a contribution increases when the contribution is prescribed by the rule, mostly at the 10% significance level. In State 1, 10 points are socially more appropriate under the Unconditional Rule (0.49 vs. 0.29, p=0.0598) while 20 points are socially more appropriate under the Conditional Rule (0.76 vs. 0.89, p=0.0556). In State 2, both rules dictate a contribution of 10 points. The social appropriateness of 10 points is the same for both rules (0.67 vs. 0.64, p=0.6447). In State 3, 10 points again are socially more appropriate under the Unconditional Rule (0.56 vs 0.33, p=0.0837) while zero points are socially more appropriate under the Conditional Rule (-0.09 vs 0.40, p= 0.0006). While negative deviations from the rule are always socially inappropriate, the evidence of positive deviations is mixed.[18]

The findings for State 2 confirm Hypothesis 2.3. Both rules dictate a contribution of 10 points for the most probable Signal II. However, a selfish deviation from the rule to a contribution of zero points is considered significantly less socially inappropriate under the Conditional Rule among rulers (-0.60 vs. -0.36, p=0.0333).

Result 2.3.
A selfish, rule-breaking contribution is socially less inappropriate under the Conditional Rule.

[18]One potential explanation is that social appropriateness includes other dimension of social behavior than rule-following, for example efficiency and social concerns.

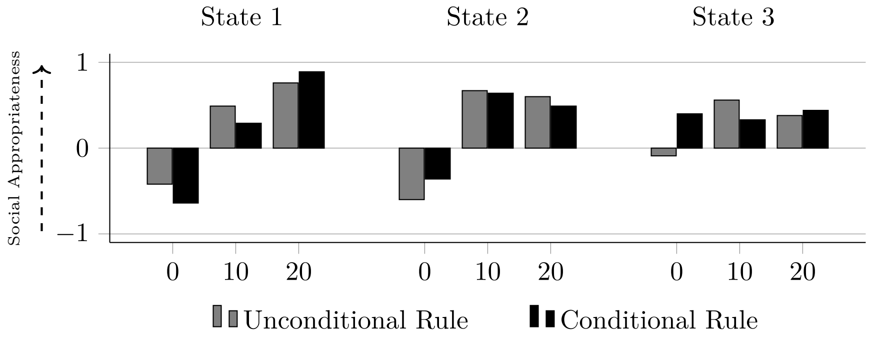


Figure 2.5: Mean Measures of Social Appropriateness of Contributions

Note: The scale ranges from -1 (“socially very inappropriate”) to 1 (“socially very appropriate”). The means are dependent on the state and the group rule.

Therefore, a neutral observer (here the rulers) takes into account that the contributor may have received a misleading signal. In this case a selfish contribution is rule-breaking only under the Unconditional Rule; hence a wiggle room emerges. Selfish subjects can use a lack of punishment (in terms of social disapproval) under the Conditional Rule to act more selfishly.

However, Figure 2.4 already reveals that subjects do not contribute less under the Conditional Rule for Signal II. The mean contributions for Signal II are not significantly different, neither in Part II (8.10 points under the Unconditional Rule and 8.05 under the Conditional Rule, p=0.9722) nor in Part III (7.97 points under the Unconditional Rule and 7.71 under the Conditional Rule, p=0.4679). Therefore, Hypothesis 2.4 has to be rejected.

Result 2.4.
Subjects do not use the wiggle room that the Conditional Rule offers them.

For Part III, additional regression analyses displayed in Table 2.4 give further insights.[19] First, in line with previous findings on repeated public goods games (Chaudhuri 2011), contributions decline over time.[20] Secondly, individuals react to the information provided by signals and adapt their contribution behavior. Signal I and Signal II increase expected earnings from a

[19]Other estimation models such as a Pooled OLS model and a Fixed Effects GLS model yield similar results and are displayed in Appendix B.1.

[20]For an illustration, see Figure B.3 in Appendix B.1.

Dependent variable: Indv. Contribution	Model 1	Model 2	Model 3	Model 4
Round	-0.35***	-0.34***	-0.35***	-0.34***
	(0.077)	(0.770)	(0.067)	(0.067)
Conditional Rule	0.94	0.74		
	(0.745)	(0.777)		
Signal I	8.70***	8.84***	6.08***	6.13***
	(0.772)	(0.782)	(0.808)	(0.813)
Signal II	6.56***	6.74***	6.16***	6.21***
	(0.358)	(0.354)	(0.468)	(0.463)
Signal I * Cond. Rule			3.98***	3.83***
			(1.423)	(1.311)
Signal II * Cond. Rule			-0.48	-0.57
			(0.654)	(0.745)
Signal III * Cond. Rule			-1.23***	-1.58**
			(0.457)	(0.734)
Avr. Contr. Part I		0.37***		0.38***
		(0.059)		(0.059)
Constant	2.71***	-2.38	3.83***	-1.02
	(0.646)	(1.541)	(0.450)	(1.519)
Controls		Yes		Yes
Number of obs.	1,620	1,590	1,620	1,590
Number of indv.	162	159	162	159
Number of group clusters	54	54	54	54

Table 2.4: GLS Random Effects Model on Contributions in Part III

Note: Robust standard errors in brackets, clustered at the group level. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.001$. Three groups were dropped because of a Rule B-group match or because of no match in the session with 18 subjects. Controls in Models 2 & 4 include age, gender, previous participations in experiments, self-reported math skills and measures for altruism, negative reciprocity, risk and trust. The question on gender was not answered by three participants, they were dropped in Models 2 & 4.

contribution compared to Signal III and consistently increase contributions as well.

In Models 1 & 2 the Conditional Rule dummy indicates no general effect of rules on contribution behavior. However the Conditional Rule has signal-specific effects as it induces higher (lower) contributions for Signal I (Signal III) as shown in Models 3 & 4. Yet, as already stated in Result 2.4, there is no difference in contributions across rules when the signal is Signal II. Generally, this finding raises additional questions. In the experiment, subjects follow rules and, according to the social appropriateness measure, there is room to act more selfishly under the Conditional Rule. However, the moral wiggle room has no effect on overall contribution behavior.

2.3.5 Additional Results

Contribution Behavior in Part I and Types. In Part I, subjects contribute to the group account for each state separately. This gives data on the subjects' behavior in "regular" public goods games. I can use these data (a) to classify subjects and examine whether rules have different impact for different subjects and (b) as an additional control for subjects' understanding of the experiment.[21] Due to the complex design, it is important to verify that subjects behave consistently across experimental parts.

First, we can establish that subjects (1) have understood the PGG's incentive structure and (2) behave consistently across parts. Subjects follow the incentives to cooperate in each state of Part I as mean contributions decline with the state.[22] Subjects also behave consistently as a subject's contributions in Part I has a significant effect on his/her contributions in Part III (see Table 2.4).[23]

Secondly, I classify subjects into "selfish subjects", "perfect contributors" and "imperfect contributors" according to their Part I behavior.[24] There are 26 selfish subjects (12.6%), 86 imperfect contributors (41.5%) and 62 perfect contributors (30.0%) who account for 84% of all subjects. As a next step, I examine whether the Unconditional Rule and the Conditional Rule have different effects on the three types' behavior in Part III. The type-specific mean contributions are displayed in Figure 2.6 for each signal. Contributions

[21] Further, 14 control questions were asked to examine subjects' understanding of the PGG and the uncertainty and group rule in Part II. Over 90% of all participants answered all questions correctly after two attempts at most.

[22] Mean contributions are 10.14 points in State 1, 8.02 points in State 2 and 1.64 points in State 3. All means differ significantly at the 1%-level, using a Wilcoxon signed-rank test.

[23] As a summary statistics for Part I, a subject's average contribution across all states was used in the regression analysis. In Table 2.4, the variable is called "Avr. Contr. Part I".

[24] Because of a lack of data on conditional contributions, I use a different classification than the standard approach by Fischbacher et al. (2001). Instead the subjects' conditional contributions, I use their contributions across states. A subject is "selfish" if he contributes zero points in all three states. A subject is a "perfect contributor" if he contributes the efficient amount in each state. A subject is an "imperfect contributor" if he contributes the efficient amount in State 1 or State 2, but less otherwise.

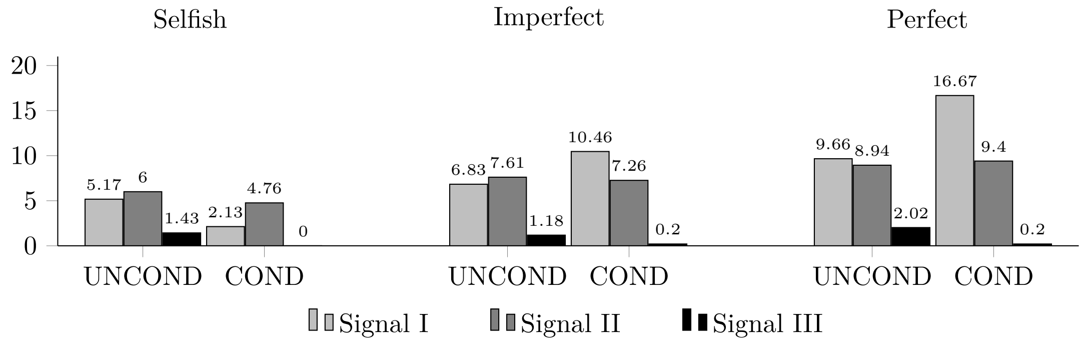


Figure 2.6: Type-Specific Mean Contributions in Part III

differ by type as expected. Perfect contributors contribute more than other types and follow both rules the most. Imperfect contributors also follow both rules, but to a lesser extent. Selfish subjects contribute less than other types. Most interestingly however, they do display wiggle room behavior; they contribute less under the Conditional Rule. However, the differences are only significant for Signal I (p=0.0060), but not for Signal II (p=0.2712).[25] This may be due to stronger incentives to act selfishly, given that the Conditional Rule prescribes a contribution of 20 points for Signal I but of 10 points for Signal II.

Result 2.5.
Selfish subjects use a moral wiggle room under the Conditional Rule if incentives to act selfishly are sufficiently strong.

Beliefs. Subjects believe that the other group members follow the rule. Whenever a rule prescribes higher contributions, subjects' beliefs follow.[26] Together with the results on social appropriateness, this is suggestive evidence for an "expressive function" of rules (as in Sunstein 1996). Here rules have no direct power as there are no formal sanctions on non-compliance. However they alter beliefs and social norms and thus change behavior.

[25]The difference for Signal III is also significant (p=0.0318), but not indicative of wiggle room behavior.

[26]See Figure B.1 in Appendix B.1. In both parts, mean belief differences are significant at the 1%-level for Signal I and Signal III but are insignificant for Signal II.

Payoffs. The experimental parameters were ex-ante constructed such that, if all subjects followed rules, i.e. without a wiggle room equilibrium, there were no payoff differences between the two rules. Result 2.4 shows no evidence of wiggle room behavior in the experiment. Consequently, there are no significant payoff differences between the two rules. In Part II, the average payoff is 26.45 (27.30) points under the Unconditional (Conditional) Rule.[27] In Part III, average payoffs are 25.34 points and 25.65 points respectively.

Order Effects. In half of the session, the order in which states were presented to the subjects was reversed. Interestingly, strong order effects arise as contributions are significantly higher for the first state that is introduced. Secondly, the rules have different effects, dependent on the order. Again, subjects seem to follow rules more strongly in their first contribution. These effects were unanticipated, but could be explained by experimenter demand effects and anchoring.[28]

2.4 Conclusion

This chapter is concerned with wiggle room behavior under different rules. More specifically, I compare prosocial behavior and welfare effects under an Unconditional Rule and a Conditional Rule, both theoretically and in a laboratory experiment. The theory highlights a potential downside of the Conditional Rule, the emergence of a wiggle room. In that case, the Conditional Rule fails to induce selfish agents to follow the rule. Indeed, the model shows that when the share of selfish, image concerned individuals as well as their image concerns are sufficiently high, the Conditional Rule results in lower welfare than the Unconditional Rule.

In a laboratory experiment, I test for the effects of the two rules on contributions and resulting welfare in a public goods game. First, I find that subjects tend to follow both rules and that selfish contributions are considered less inappropriate under the Conditional Rule. Therefore, there is scope

[27] When stating their beliefs about average payoffs, rulers systematically underestimated average payoffs for both rules (22.30 points for the Unconditional Rule and 23.48 points for the Conditional Rule) but also expected no differences in payoffs (p=0.5149).

[28] A more thorough analysis as well as a discussion are relegated to Appendix B.1.

for wiggle room behavior; nonetheless I do not find generally lower contributions under the Conditional Rule. However, consistent with wiggle room behavior, selfish subjects do contribute less under the Conditional Rule if incentives to act selfishly are sufficiently strong.

Therefore, the evidence on wiggle room behavior and subsequent implications on the optimality of rules remains inconclusive. Consistent with findings by van der Weele et al. (2014), I do not find general evidence of wiggle room behavior. One potential reason is that "reciprocity trumps wiggling". Both experiments include subjects' cooperation (trust and moonlighting games in van der Weele et al. (2014), a public good game here). If reciprocity concerns are sufficiently strong they may outweigh the incentives to act selfishly, even in wiggle room situations. Consequently the Conditional Rule may be superior in similar settings. On the other hand, selfish subjects seem to behave more selfishly under the Conditional Rule when incentives are sufficiently strong, in line with wiggle room behavior found in Dana et al. (2007), Andreoni and Bernheim (2009), and Grossman (2014).

Generally, the lack of overall effects of wiggle room behavior may thus be attributed to reciprocal preferences that make the use of wiggle rooms unattractive, or inadequate incentives to act selfishly or to a surprisingly low fraction of selfish subjects. In the current experiment, only 13% of the subjects were categorized as selfish, compared to, e.g., 30% in Fischbacher et al. (2001). In conclusion, more research is needed to fully understand the drivers of wiggle room behavior in different settings.

Individuals easily find excuses in manifold situations to explain their selfish behavior (e.g. Exley 2015); the intransparency in moral wiggle rooms is only one of them. To disregard rules, more excuses come to mind. For example, Unconditional Rules may be seen as implausible or inefficient and therefore never come into effect. Conditional Rules especially lack simplicity; here individuals may use complexity as an excuse for non-compliance. Future research should further investigate what excuses are induced by different rules and how they affect the rules' effectiveness.

Chapter 3

Promotion, Delegation and Selection

3.1 Introduction

Internal promotions are widely used to fill job vacancies.[1] They are ought to achieve two objectives at once, namely to create incentives for employees and to select the best managers (Milgrom and Roberts 1992). It is not obvious, however, that the two objectives are always in line. To motivate employees, promotions must be based on current employee performance. To select the best managers, they must be based on expected manager performance. If the two measures are negatively correlated, a conflict of objectives arises.

There is ample evidence that promotions are mostly based on past and current employee performance (Frederiksen et al. 2017); and the Peter principle states that this approach leads to worse management selection, in line with a conflict of objectives (Peter and Hull 1969). That claim, a trade-off between motivation and selection, is substantiated in Benson et al. (2018). They provide evidence that firms rather promote the best performing employees than the expectedly best managers. Indeed, firms could improve managerial performance by 30% if promotion decisions were based on other measures than employee performance.

[1]For example, DeVaro et al. (2019) find that around 60% of new jobs are filled via internal promotions. Baker et al. (1988, p.600) state that “promotions are used as the primary incentive device in most organizations”. For a broader overview on empirical studies, see Gibbons and Roberts (2012).

This chapter examines an arising trade-off between the two objectives of performance-based promotions, motivation and selection. Most importantly, it first illustrates that such trade-off emerges from misaligned incentives between managers and the firm and the delegation of decision rights. That approach differs from previous "skill-based" models such as Fairburn and Malcomson (2001), Lazear (2004), and Schöttner and Thiele (2010) and from the original intuition behind the Peter principle (Peter and Hull 1969). These models blame skill differences between employee and manager tasks for bad manager selection of performance-based promotions. In contrast, my theory offers another explanation, namely the exploitation of decision rights by promoted managers.

I show that employees who generate lower profits as managers are *incentivized more strongly* by the prospect of becoming manager; consequently they are also more likely to get promoted. The intuition stems from the fact that a management position, i.e. the promotion prize, entails more responsibility and more decision rights. Yet managerial decision-making is not contractible. Hence a manager can exploit decision rights at his own interest; only the preference alignment between manager and principal determines how private benefits and profits are influenced by the manager's decisions. Consequently, lower preference alignment leads to higher private benefits for the manager and lower profits for the firm.

It follows that those employees who gain more from a promotion because of higher private benefits generate lower profits. If promotions are based on employee performance, a trade-off between motivation and selection arises. First, employees work hard to get promoted as they are incentivized by future private benefits as managers. Furthermore, employees with lower preference alignment receive higher private benefits when promoted and thus they will work even harder than their competitors. Such behavior leads to a higher promotion probability and a negative selection effect arises: employees who generate lower profits as managers are more likely to get promoted.

The model further investigates the trade-off between motivation and selection from different angles. First it examines how a principal should optimally delegate decision rights when promotions are based on employee performance. Since managers receive private benefits from decision-making, the principal could limit their decision rights, and thus private benefits, by delegating fewer decisions. As a result, bad selection of managers is reduced at the expense of employee motivation. Consequently the optimal level of delegation balances both effects. One of the main findings shows that partial delegation can be optimal under performance-based promotions: the principal delegates only a limited number of decisions to the manager as too much delegation would attract unprofitable managers too strongly.

Moreover, I analyze the optimality of performance-based promotions in light of the manager's decision rights. If a management position is equipped with many decision rights, the selection effect will be particularly pronounced. In this case the principal will not use performance-based promotions to select managers, to prevent those with worse preference misalignment from rising to the top. I further consider the optimal design of the organization, i.e. the joint choice of delegation and promotion. I show that (a) delegation and the simultaneous use of performance-based promotions lead to higher motivation of employees and (b) their joint use increases if preferences are sufficiently aligned and the selection effect is sufficiently low.

The model also offers a new perspective to why we observe promotions at all when the Peter principle applies. Fairburn and Malcomson (2001, p.45) ask: "Why not (..) use promotions to assign employees to jobs and monetary bonuses to provide incentives?" I investigate this question in two extensions. I first analyze how worker wages and delegation interact. Workers are motivated by both wages and expected private benefits from delegation, implying a substitutability between the two incentive devices. Higher managerial discretion then comes with lower subordinates' wages. Overall, including monetary incentives decreases the number of decisions delegated, but, in general, partial delegation remains optimal. Secondly, I introduce further promotion-related wage increases (Baker et al. 1994). I show that if such a wage increase is optimal for the principal, the joint use of performance-based promotions and delegation of decision rights is profitable as well.

Related Literature. This work combines two distinct strands of literature, namely that on the (negative) selection effects of promotions and the optimal delegation of decision rights. First, it offers a new explanation for why performance-based promotion schemes induce inefficient selection of managers.[2] While theories differ in explaining the benefits of performance-based promotions[3], they share the intuition behind its inefficiency: performance-based promotions result in managers with an insufficient skills. Accordingly, low manager performance arises because promoted workers lack the required skills to be good managers (Peter and Hull 1969; Bernhardt 1995; Fairburn and Malcomson 2001; Lazear 2004; Schöttner and Thiele 2010; Koch and Nafziger 2012). The current approach is different as it is fully independent of skill considerations. In my model, low manager performance arises because promoted workers exploit managerial decision rights; and they vary in the degree that they do so.

I also contribute to the literature on optimal delegation in which the principal delegates decision rights to the manager to make him, e.g., acquire information (Aghion and Tirole 1997), communicate truthfully (Dessein 2002), or exert effort (Bester and Krähmer 2008). In contrast to these papers, I am concerned with the effects of delegation on the behavior of *employees*. My model focuses on the link between managerial benefits and employees' behavior via performance-based promotions. It is silent about potential sources of these benefits as well as the manager's decision problem. Empirical evidence regarding managerial private benefits can be, e.g., found in the literature on managerial empire building (Jensen 1986; Hope and Thomas 2008), short-termism (Bebchuk and Stole 1993; Edmans et al. 2017) and intrinsic valuation of decision rights (Fehr et al. 2013; Bartling et al. 2014). In general, managerial private benefits can arise from misaligned preferences as well as ill-designed incentive schemes (see the discussion in Dessein 2002, p.815).

Furthermore, this chapter connects to work on complementarities in organizational design (see Milgrom and Roberts 1990; Roberts 2007). There often exist interaction effects between different dimensions of organizational design, for example between job design, monitoring and incentives (so called

[2]Empirical evidence for that claim can be found in Grabner and Moers (2013) and Benson et al. (2018). They show promoting high-performing employees correlates negatively with manager quality. More specifically, Benson et al. (2018) find that firms could increase management performance by 30% if they based promotions on other measures than employee performance.

[3]For example, in Fairburn and Malcomson (2001) they prevent influence activities. In my model, in Schöttner and Thiele (2010) and Koch and Nafziger (2012) and a large literature on tournaments (Rosen 1986), performance-based promotions are used because they act as an incentive device.

"high performance work systems", see Ichniowski and Shaw 2003), or between the hierarchical structure of an organization and its use of promotions (Ke et al. 2018). In the current model, complementarities between delegation and the use of promotions arise because delegation increases private benefits of managers and thus the prize for winning a promotion.

Lastly, my work complements a literature in political economy on the delegation of authority and selection effects. Already Knight (1938) and Hayek (1944) discussed the influence of institutions and political systems on the selection of politicians. Besley (2005) points out that a political office's "attractiveness" to candidates crucially depends on the rents they can extract while in office; these in turn depend on the office's power. Such consideration will affect who is running for office, "egoistic" or "public-spirited" politicians.[4] Similarly, the current model shows that decision rights, the "attractiveness" of a management position, attract employees who want to exploit those decision rights. Therefore, rents to promotion and power must be limited to mitigate potential selection effects. For political offices, this can be done, e.g., by setting up institutions to align a politician's private interests (Barro 1973), his accountability, for instance via re-elections (Maskin and Tirole 2004; Acemoglu et al. 2010), the implementation of "checks and balances" (Persson et al. 1997), or power de-concentration (Grunewald et al. 2019). In organizations, the principal can simply restrict a manager's decision rights.

The remainder of the chapter is as follows. In Section 3.2 I introduce the model and then analyze optimal delegation and optimal promotion rules in Section 3.3. Section 3.4 presents two extensions of the model by introducing monetary incentives for the workers, via bonus schemes and via promotion-related wage increases. Section 3.5 concludes.

3.2 Model

Overview

The firm consists of a principal (female) and two workers (male) who exert effort and compete for a promotion to a management position. The principal

[4]This idea of political selection is prevalent even in science-fiction novels. To quote David Brin, the author of the post-apocalyptic novel "The Postman" (1985, p. 267): "It is said that power corrupts, but actually it's more true that power attracts the corruptible."

receives profits from both hierarchy levels, i.e. the workers' work effort and from the decisions made by the manager who is a promoted worker. In order to maximize profits, the principal ex-ante chooses an organizational design that has two dimensions. Thereby he chooses a promotion rule and the degree of delegation of decision rights to the management position.

The model is introduced step-by-step. First I present the workers' effort choice and how it is shaped by promotion prospects. Then I continue by introducing the management stage and the delegation decision. The model setup is concluded by bringing both parts together. The incorporation of wage payments, to the manager or to the employees, is relegated to the model extensions in Section 3.4. Section 3.5 concludes.

Promotions and Workers' Effort

In the firm there are two workers competing for a promotion to a management position. A worker i exerts unobservable effort $e_i \in [0, 1]$ on a project at convex effort costs $c(e_i) = \frac{c}{2}e_i^2$. Each worker's project can either be a success or a failure. In case of success, the principal receives $S > 0$, otherwise 0. A worker i's project success probability, given by $pr(success) = e_i$, increases linearly in his effort and is independent from the other worker's effort.

Workers are incentivized purely by promotion prospects and do not receive any wage payments.[5] The principal P (she) ex-ante commits to a promotion rule. The promotion rule is fully captured by a promotion probability $p_i(e_i, e_j)$ for a worker i, given i's and his coworker j's effort levels e_i and e_j.

If worker i gets promoted he receives private benefits as a manager. These are denoted by u_{m_i} and will depend on the delegation decision as introduced later. In general the risk-neutral worker i's utility function is given by

$$u_i(e_i) = p_i(e_i, e_j) \cdot u_{m_i} - \frac{c}{2}e_i^2. \tag{3.1}$$

Promotion Rules

The principal can decide between two promotion rules $prom \in \{\mathcal{P}, \mathcal{R}\}$. The "performance-based promotion" $\mathcal{P}$ is based on the workers' project outcomes.

[5] In Section 3.4.1, workers receive additional performance-based wage payments. The main results remain unchanged.

When only one project is successful the principal promotes the successful worker. Otherwise, he randomizes between the workers.[6] In contrast, the "random promotion" $\mathcal{R}$ is fully independent of the workers' work. In that case, the principal randomizes between the workers and chooses each of them with probability $p^{\mathcal{R}} = 0.5$.[7]

Delegation

In the firm, a finite divisible number of similar decisions, normalized to 1, need to be made. The principal P can delegate $k \in [0,1]$ of these decisions to a manager M (he) who can then implement his favored choice. The payoffs for each decision depend on the decision-maker:

1. If P makes a decision, total surplus from this decision is $\pi > 0$. The manager cannot extract any private benefits and thus his payoff is $u_M = 0$. The principal receives the full surplus, thus her profits are $\Pi = \pi$.
2. If M makes a decision, total surplus from this decision is $\pi^D > \pi$. But the manager extracts a share $\alpha \in [0,1]$ of the surplus, and his payoff is $u_M = \alpha\pi^D$. The principal receives the remaining share, $\Pi = (1-\alpha)\pi^D$.

The principal delegates decision rights over a fraction of k decisions to the manager, and keeps decision rights over a fraction of $(1-k)$ decisions to herself. Thus overall payoffs from managerial decision-making are given by

$$\Pi(k) = k \cdot (1-\alpha)\pi^D + (1-k) \cdot \pi = \pi + k \cdot (\delta - \alpha)\pi^D \tag{3.2}$$

$$u_M(k) = k \cdot \alpha\pi^D. \tag{3.3}$$

Here, $\delta = \frac{\pi^D - \pi}{\pi^D} \in (0,1)$ displays the relative surplus increase due to better managerial decision-making. The setup resembles the main trade-off of a standard delegation problem in a stylized way.[8] A manager makes overall

[6] In Appendix C.1, I show that these two promotion rules are superior to any convex combination of the two. Therefore, if it is optimal not to randomize fully between the workers, it is optimal not to randomize at all.

[7] One famous example for purely random promotions in a slightly different setting is that of ancient Athens. There, political offices were filled via lots to ensure "pure" democracy (see Headlam 1891). I thank Mike Powell for bringing up this example.

[8] For example, it arises from Aghion and Tirole (1997) with the following parameter values: $\alpha^{AT} = (1-\alpha)\frac{\pi^D}{\pi}, \beta^{AT} = 0, b^{AT} = \pi^D, B^{AT} = \pi$, effort levels of $e^{AT} = E^{AT} = 1$ and normalized costs of $g_P^{AT}(1) = g_A^{AT}(1) = 0$.

better decisions than the principal, captured by $\pi^D > \pi$.[9] However, delegation also comes with a loss of control which stems from the decisions' non-contractibility and a preference misalignment between manager and principal. The degree of preference misalignment is measured by α. Increasing α implies a stronger preference misalignment. Profits fall and private benefits rise.

Unknown Managerial Types

There are two different types of managers that have distinct degrees of preference misalignment. Manager types are private information. Each worker i is one of the two manager types, denoted by $m_i \in \{A, B\}$, with $prob(m_i = B) = \mu \in (0, 1)$. Both types are equally skilled and generate the same total surplus when making a decision, given by π^D. However, the types' preference alignment differs. A-type managers' preferences are well aligned, B-type managers' are not. I assume that $\alpha_B > \delta > \alpha_A$. Thus the principal would delegate all decisions to an A-type, but none to a B-type manager *if she knew the type.*

From a worker's perspective, a promotion hence yields private benefits of $u_{m_i}(k)$ that depend on his private type m_i as well as the management position's amount of decision rights k. Therefore, worker i's private benefits from a promotion are given by $u_{m_i} = k\alpha_{m_i}\pi^D$. I can re-write worker i's utility function as[10]

$$u_i(e_i) = p_i \cdot \underbrace{k \cdot \alpha_{m_i}\pi^D}_{u_{m_i}} - \frac{c}{2}e_i^2. \tag{3.4}$$

Timeline

To conclude the model setup, the time structure is as follows.

1. The principal chooses the degree of delegation k and the promotion rule *prom*.

2. Workers are independently drawn from the population with respect to their managerial type.

[9]Reasons for the superiority of managers' decision-making include local information that are available only to the manager (Hayek 1945) or that delegation increases the manager's initiative (Aghion and Tirole 1997).

[10]In the following, whenever possible, I drop the subscript i for notational ease.

3. Workers simultaneously exert unobservable effort, and each worker's project outcome is realized and observed.

4. The principal promotes one of the workers according to the promotion rule chosen in $t = 1$. The other worker leaves the firm and receives an outside utility of 0.

5. Decision rights are delegated to the newly promoted manager, according to the choice in $t = 1$. Payoffs from decision-making are realized and the game ends.

3.3 Analysis

In this section I analyze the model presented above. First I examine the optimal choice of delegation for each promotion rule. Then I derive the optimal promotion rule, having fixed the degree of delegation. At the end of this section I analyze the optimal joint decision of delegation and promotion rule. All proofs can be found in Appendix C.2.

3.3.1 The Effects of Promotion Rules on Delegation

Random Promotion

Random promotion implies a fixed promotion probability of $p^{\mathcal{R}} = 0.5$ for each worker that is independent of effort. A worker cannot influence the probability of promotion and thus has no incentives to work. It follows that $e^{\mathcal{R}} = 0$ for both workers. Then the principal faces a B-type manager with probability μ, as if there was a random draw from the population. This is because workers are drawn independently from the population and are also promoted randomly. Let $\overline{\alpha} = (1-\mu)\alpha_A + \mu\alpha_B$ denote the expected preference misalignment. As $e^{\mathcal{R}} = 0$, projects are unsuccessful with certainty and the principal's profits only consist of the payoff from decision-making:

$$E\Pi^{\mathcal{R}} = \mu\Pi_B + (1-\mu)\Pi_A = \pi + k(\delta - \overline{\alpha})\pi^D \tag{3.5}$$

where $\Pi_m = \pi + (\delta - \alpha_m)\pi^D$ are decision-making profits when facing manager type $m \in \{A, B\}$.

Consequently, optimal delegation under random promotion depends on the relative benefits from managerial decision-making and losses from *expected* preference misalignment. Note that the principal's profits are linear in k. Therefore, if and only if the benefits from managerial decision-making, given by δ, outweigh the expected loss of control due to preference misalignment, $\overline{\alpha}$, the principal delegates all decision rights to the manager, and none otherwise. This is summarized in Proposition 3.1.

Proposition 3.1.
Under random promotion, the principal either delegates all decisions, or none:

$$k^{\mathcal{R}} = \begin{cases} 1 & \text{if } \delta \geq \overline{\alpha} \\ 0 & \text{if } \delta < \overline{\alpha}. \end{cases} \tag{3.6}$$

Performance-Based Promotion

Under performance-based promotions each worker's promotion probability depends on the respective project success. It follows that workers face a strategic game in which their expected utility and thus their optimal strategy depend both on their own and their co-worker's exerted effort. However, workers observe neither their co-worker's managerial type nor their effort level. This game is simplified by its binary structure. A worker i's expected promotion probability is given by

$$\begin{aligned} p_i =& \mu \underbrace{\left[\underbrace{e_i(1-e_{B_j})}_{i \text{ is successful}} + \underbrace{0.5 e_i e_{B_j}}_{\text{both successful}} + \underbrace{0.5(1-e_i)(1-e_{B_j})}_{\text{both unsuccessful}}\right]}_{\text{The other worker is a B-type}} \\ &+ (1-\mu) \underbrace{\left[\underbrace{e_i(1-e_{A_j})}_{i \text{ is successful}} + \underbrace{0.5 e_i e_{A_j}}_{\text{both successful}} + \underbrace{0.5(1-e_i)(1-e_{A_j})}_{\text{both unsuccessful}}\right]}_{\text{The other worker is a A-type}} \end{aligned} \tag{3.7}$$

which can be simplified to

$$p_i = 0.5 + 0.5(e_i - \overline{e}_j). \tag{3.8}$$

Here, $\overline{e}_j = (1-\mu)e_{A_j} + \mu e_{B_j}$ denotes the ex-ante expected effort level of a worker with e_{m_j} defining the effort level of worker j with managerial type m.

The resulting functional form of p_i is a Difference Contest Success Function with unknown contenders (Hirshleifer 1989). It follows that the marginal effect of i's effort on his promotion probability is independent of his expected co-worker's effort. Therefore, each worker has unique dominant strategy that is derived by standard utility maximization. Lemma 3.1 states optimal effort provision in the resulting Nash equilibrium.

Lemma 3.1.
A worker's optimal equilibrium effort under performance-based promotions increases in the degree of delegation and is higher for B-type workers and is given by

$$e_m^{\mathcal{P}} = k \cdot \frac{\alpha_m \pi^D}{2c}. \tag{3.9}$$

Workers exert effort, i.e. $e_m^{\mathcal{P}} > 0$, only if there is a positive amount of delegation, i.e. $k > 0$. Moreover, because private benefits of a manager are increasing in both the degree of delegation and the preference misalignment, so are expected utility and effort provision. This is the *incentive effect* of performance-based promotions. Moreover, since the preference misalignment of a B-type manager is larger, i.e. $\alpha_B > \alpha_A$, B-type workers exert higher effort than A-types. This translates into a higher probability of promotion for the B-type. A *selection effect* arises, stated in Lemma 3.2a. Note that the selection effect only arises with a heterogeneous workforce, i.e. if there are one A-type as well as one B-type worker among the workers. In a homogeneous workforce, both workers are of the same type, exert the same effort and are promoted with the same probability, as described by Lemma 3.2b.

Lemma 3.2.
Under performance-based promotion,

(a) with a heterogeneous workforce with one A-type and one B-type worker, the following statements hold if and only if there is delegation ($k > 0$):

- *B-type workers are promoted with a higher probability than A-types:*

$$p_B^h - p_A^h = k \cdot \frac{(\alpha_B - \alpha_A)\pi^D}{2c} > 0 \tag{3.10}$$

where p_m^h denotes the probability of success for type m in a heterogeneous workforce.

- *The types' difference in their probability of promotion is increasing in the degree of delegation: $\frac{\partial(p_B^h - p_A^h)}{\partial k} > 0$.*

(b) with a homogeneous workforce, workers exert identical, but type-dependent effort and thus have the same probability of promotion of $p_m^{hom} = 0.5$.

Lemma 3.2a says that workers who make less profitable decisions are promoted with a higher probability. This is a mildly revised, more general version of the Peter principle as it implies that, in expectations, profits are reduced by promoting the worse manager. The mechanism in Lemma 3.2 also implies that a negative selection effect persists even when the principal bases the promotion decision on other measures than current performance, provided that workers still can influence their promotion probability, for instance by shifting effort towards the new promotion measure or by gaming.

Taking optimal worker behavior from Lemmas 3.1 and 3.2 as given, the principal maximizes his own expected profits $E\Pi^{\mathcal{P}}$ over the degree of delegation k. Since managerial types are private, expected profits are given by

$$E\Pi^{\mathcal{P}} = \overbrace{(1-\mu)^2[\Pi_A + 2e_A^{\mathcal{P}}S]}^{\text{only A-type workers}} + \overbrace{\mu(1-\mu)[p_A^h\Pi_A + p_B^h\Pi_B + 2\overline{e}^{\mathcal{P}}S]}^{\text{heterogenous workforce}} + \overbrace{\mu^2[\Pi_B + 2e_B^{\mathcal{P}}S]}^{\text{only B-type workers}} \tag{3.11}$$

with $\overline{e}^{\mathcal{P}} = (1-\mu)e_A^{\mathcal{P}} + \mu e_B^{\mathcal{P}}$ and optimal effort levels given by Lemma 3.1. Expected profits can be displayed by considering each potential workforce composition. For each of these, profits depend on the (expected) payoff from managerial decision-making and workers' project success. Moreover, Equation (3.11) can be decomposed into three parts:

$$\begin{aligned} E\Pi^{\mathcal{P}} &= \underbrace{(1-\mu)\Pi_A + \mu\Pi_B +}_{} \quad \underbrace{2\cdot\overline{e}^{\mathcal{P}}S}_{} \quad + \underbrace{2\mu(1-\mu)\frac{(p_B^h - p_A^h)(\Pi_B - \Pi_A)}{2}}_{} \\ &= \overbrace{\pi + k(\delta - \overline{\alpha})\pi^D +}^{\text{Random Promotion}} \quad \overbrace{\frac{k\overline{\alpha}\pi^D}{c}S}^{\text{Incentive Effect}} \quad \overbrace{-\,\mu(1-\mu)\frac{(k(\alpha_B - \alpha_A)\pi^D)^2}{2c}}^{\text{Selection Effect}}. \end{aligned} \tag{3.12}$$

Compared to random promotions, a performance-based promotion induces two further effects on expected profits. It increases worker incentives and thus expected gains from successful projects (the incentive effect). On the other hand, it worsens management selection and lowers expected profits from managerial decision-making by promoting the "wrong kind of manager" with a higher probability (the selection effect). Profit maximization leads to the optimal degree of delegation, given in Proposition 3.2.

Proposition 3.2.
Under performance-based promotion, optimal delegation is given by

$$k^{\mathcal{P}} = \begin{cases} 1 & \textit{if } \delta \geq \alpha_2 \\ \tilde{k} & \textit{if } \delta \in [\alpha_1, \alpha_2) \\ 0 & \textit{if } \delta < \alpha_1 \end{cases} \tag{3.13}$$

with $\tilde{k} = \frac{c(\delta-\overline{\alpha})+\overline{\alpha}S}{\mu(1-\mu)(\alpha_B-\alpha_A)^2\pi^D}$, $\alpha_1 = (1-\frac{S}{c})\cdot\overline{\alpha}$ *and* $\alpha_2 = (1-\frac{S}{c})\cdot\overline{\alpha}+\frac{\mu(1-\mu)(\alpha_B-\alpha_A)^2\pi^D}{c}$.

Figure 3.1 gives a graphical illustration of that result. One can explain optimal delegation by focusing on the three effects of delegation on expected profits, as displayed in Equation (3.12). First, optimal delegation for random promotion is given by Proposition 3.1. If $\delta \geq \overline{\alpha}$, there is full delegation, and none otherwise. This is depicted in Figure 3.1 as the dotted line. Under performance-based promotions, two additional effects come into play. The incentive effect on worker behavior increases profits as more delegation makes both workers work harder and thus increases the probability of project success, as is shown in Lemma 3.1. Generally the incentive to delegate increases if the expected loss from preference misalignment is sufficiently low, i.e. $\alpha_1 < \delta$. The selection effect only arises in a heterogeneous workforce which occurs with probability $2\mu(1-\mu)$, see Lemma 3.2. It is a combination of two distinct effects as delegation affects both workers' and the manager's behavior. Since delegation affects workers' effort differently, it increases the probability that a B-type worker gets promoted, $\frac{\partial(p_B^h-p_A^h)}{\partial k} > 0$. It also increases the relative loss in profits when a B-type manager takes decisions instead ot an A-type manager, $\frac{\partial(\Pi_B-\Pi_A)}{\partial k} < 0$. Additionally, there exists an interaction between the two effects. The severity of an increase in B-type's

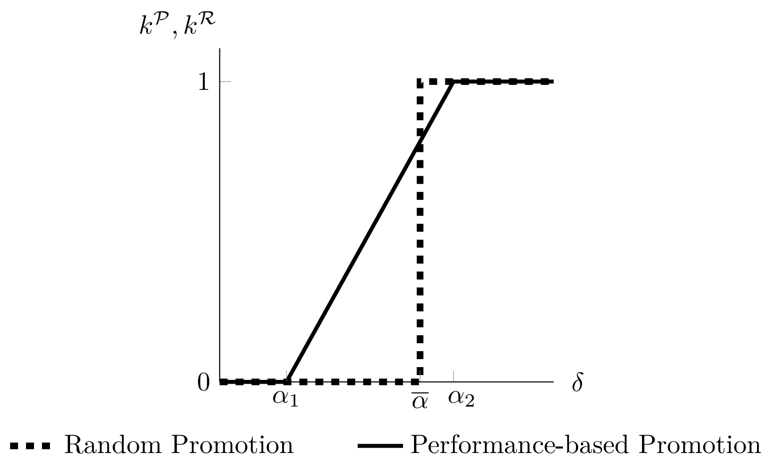


Figure 3.1: Optimal Delegation under Random and Performance-Based Promotions

probability of getting promoted depends on the loss that is related to B-type managers, and vice versa.

To gain more intuition, suppose $k = 0$. Then, the relative loss and the difference in promotion probabilities are zero as well. Increase k marginally. The effect of a marginal increase on profits is given by the marginal increase in B-type promotion probability, *holding fixed the relative loss*, and the marginal increases in the relative loss, *holding fixed promotion probabilities.* A marginal increase in B-type promotion probability does not affect profits as the relative loss at $k = 0$ is still zero, and vice versa. Therefore, at $k = 0$, a marginal increase in delegation *does not affect* profits via the selection effect. On the other hand, suppose k is close to 1. Then, a marginal increase in k is severe as (a) the marginal effect on promotion probabilities is large because relative loss is high already and (b) the marginal effect on the relative loss matters because the difference in promotion probabilities is also high. Thus, even though profits from decision-making and optimal effort level are linearly increasing in k, expected profits are quadratic in k, due to the selection effect. This intuition is summarized in Lemma 3.3.

Lemma 3.3.
The selection effect of delegation under performance-based promotion is an increasing, quadratic function of k, being zero at $k = 0$. It decreases expected profits for any $k > 0$.

There are two implications of Lemma 3.3. First, because the selection effect does not exist at $k = 0$, the threshold α_1 is unaffected by selection concerns. Secondly, under a performance-based promotion the optimal degree of delegation is continuous for intermediate δ, as shown in Proposition 3.2. While the incentive effect makes delegation profitable for $\delta \in (\alpha, \overline{\alpha})$, the convexity of the selection effect makes full delegation too costly if $\delta \in [\overline{\alpha}, \alpha_2)$.

Comparative Statics. If the manager's relative advantage in decision-making is small (low δ) or the expected loss of control is large (high $\overline{\alpha}$), it is not worthwhile for the principal to delegate because profits from delegation will be low. The same applies if gains from worker incentivization are sufficiently small because of a low value of successful projects (low S) or very high effort costs (high c).

On the other hand, delegation is always positive if projects are sufficiently profitable, i.e. if $S > c$. In this case, the incentive effect is positive and outweighs potential losses due to worse decision-making.[11] Furthermore, for sufficiently high profits from managerial decision-making, the principal delegates all decisions if the project implications from bad selection are sufficiently harmless, for instance because a heterogeneous workforce is unlikely (μ close to 0 or 1). Selection is also a minor concern if the heterogeneity among managers is low (low $\alpha_B - \alpha_A$).

Partial delegation arises whenever gains from worker incentives are sufficiently high ($\delta > \alpha_1$) but the selection effect is sufficiently strong ($\delta < \alpha_2$).

Does performance-based promotion always lead to higher delegation compared to random promotion? The answer again depends on the relative strength of the selection effect. It determines the size of α_2, the "full delegation" threshold under performance-based promotion. If the selection effect is sufficiently strong, then $\alpha_2 > \overline{\alpha}$ and the principal uses partial delegation under performance-based promotion and full delegation under random promotion for $\delta \in [\overline{\alpha}, \alpha_2]$. However, if the selection effect is relatively weak compared to the incentive effect then optimal delegation is always higher under performance-based promotion. This finding is summarized in Corollary 3.1 and depicted in Figure 3.2.

[11] One can see that as, at $k = 0$, the marginal effect of increasing k on expected profits is given by $(\delta - \overline{\alpha} + \frac{\overline{\alpha}S}{c})\pi^D$ which is positive for $S > c$.

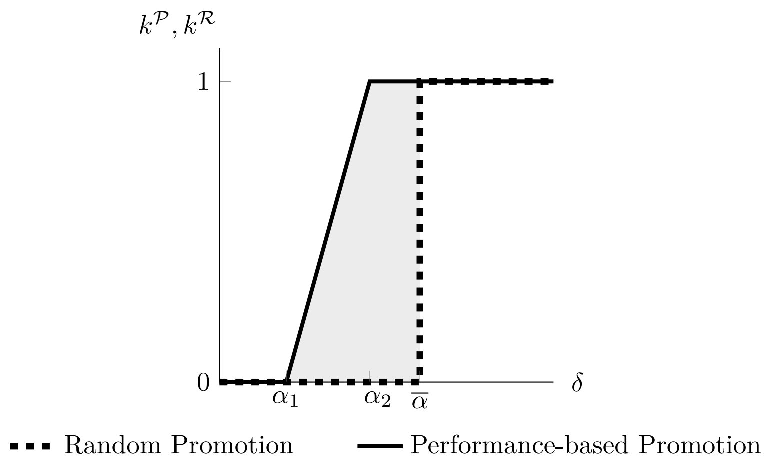


Figure 3.2: Higher Delegation under Performance-Based Promotions

Note: This is an illustration of Corollary 3.1. If $\alpha_2 < \overline{\alpha}$, delegation is always weakly higher under performance-based promotion, as displayed by the gray area.

Corollary 3.1.
Performance-based promotions induce higher delegation than random promotion if $\mu(1-\mu)(\alpha_B - \alpha_A)^2\pi^D < \overline{\alpha}S$*, i.e. if the selection effect is sufficiently weak.*

Corollary 3.1 shows a positive correlation between the use of performance-based promotions and the degree of delegation for a sufficiently minor selection effect. This is consistent with empirical evidence provided by Alfaro et al. (2016). They find a positive correlation between a plant's incentive practices, such as performance-based promotions, and the degree of delegation from headquarters to that plant.[12]

3.3.2 The Effects of Delegation on Promotion Rules

Following Corollary 1, I further analyze the optimal choice of promotion rule holding fixed the degree of delegation. That mimics manifold situations in which the principal is bound to a positive degree of delegation, at least in the short term. Reasons include information overload, time constraints and hierarchical structures within the firm. In the model, I fix k at $\overline{k} > 0$. The

[12]Thereby, they use data on management policies from the World Management Survey. Its index on incentive practices includes two measures on the use of performance-based promotions, bonuses and talent management.

principal can only decide on the promotion rule. Comparing the expected profits under the two promotions practices, gives by Equations (3.5) and (3.12), implies that the optimal promotion practice ultimately depends on the incentive and the selection effect:

$$E\Pi^{\mathcal{P}} - E\Pi^{\mathcal{R}} = \underbrace{2 \cdot \overline{e}^{\mathcal{P}} S}_{\text{Incentive Effect}} + \underbrace{2\mu(1-\mu)\frac{(p_B^h - p_A^h)(\Pi_B - \Pi_A)}{2}}_{\text{Selection Effect}}. \tag{3.14}$$

Thus, performance-based promotions are optimal if and only if the selection effect is sufficiently small. This in turn depends on the degree of delegation. The selection effect leads to a disproportionate reduction in profits with increasing delegation. Therefore performance-based promotions outperform random promotions for sufficiently low levels of $\overline{k}$, as stated in Corollary 3.2.

Corollary 3.2.
For a fixed degree of delegation $\overline{k}$, performance-based promotions outperform random promotions if and only if $\overline{k} < \overline{k}^{\mathcal{P}} = \frac{2\overline{\alpha}S}{\mu(1-\mu)(\alpha_B-\alpha_A)^2\pi^D}$, i.e. if the degree of delegation is sufficiently low .

Note that $\overline{k}^{\mathcal{P}}$ is decreasing in the size of the selection effect, $\mu(1-\mu)(\alpha_B - \alpha_A)^2\pi^D$, and increasing in the expected profits from motivating workers, $2\overline{\alpha}S$. Furthermore, performance-based promotion is optimal for any degree of delegation if the selection effect is sufficiently weak because in that case $\overline{k}^{\mathcal{P}} \geq 1$. If the selection effect is strong, and the degree of delegation is sufficiently high ($\overline{k} > \overline{k}^{\mathcal{P}}$), the principal may refrain from performance-based promotion. This is related to a finding by Benson et al. (2018). They show that the use of performance-based promotion is decreasing in the manager's team size. Under the assumption of a positive correlation between team size and the team manager's decision rights, Corollary 3.2 provides an explanation. The negative selection effect of performance-based promotions outweighs the expected benefits from worker motivation.

3.3.3 Optimal Organizational Design

When designing an organization, the principal must jointly consider all dimensions of organizational design. Only then she accounts for potential complementarities between different dimensions of organizational design as argued by Roberts (2007). Otherwise, the implementation of a certain policy on one dimension can well be ineffective, or even harmful, without complementary changes in another.[13] In the current model the principal jointly decides over the optimal degree of delegation and the promotion rule, i.e. she chooses $(k^*, prom^*) \in \{(k^{\mathcal{P}}, \mathcal{P}); (k^{\mathcal{R}}, \mathcal{R})\}$ with

$$(k^*, prom^*) \text{ arg max } \{E\Pi^{\mathcal{P}}(k^{\mathcal{P}})\, ;\, E\Pi^{\mathcal{R}}(k^{\mathcal{R}})\}. \tag{3.15}$$

Indeed, the optimal organizational design as described in Proposition 3.3 does make use of complementarities between delegation and promotion rules. Profitability of delegation is higher under performance-based promotion practices unless the selection effect is too strong, and vice versa.

Proposition 3.3.
The optimal organizational design is as follows.

- *If $\delta \leq \alpha_1$, there is no delegation and the promotion rules are equivalent:*

$$(k^*, prom^*) = (0, \mathcal{P}) = (0, \mathcal{R}). \tag{3.16}$$

- *If $\delta \in (\alpha_1, \overline{\alpha}]$, performance-based promotion with delegation is optimal:*

$$(k^*, prom^*) = (k^{\mathcal{P}}, \mathcal{P}). \tag{3.17}$$

- *If $\delta > \overline{\alpha}$, performance-based promotion with delegation is optimal if and only if the negative selection effect is sufficiently small:*

$$(k^*, prom^*) = \begin{cases} (k^{\mathcal{P}}, \mathcal{P}) & \text{iff } \mu(1-\mu)(\alpha_B - \alpha_A)^2 \pi^D \leq \frac{[c(\delta-\overline{\alpha})+\overline{\alpha}S]^2}{c(\delta-\overline{\alpha})} \\ (1, \mathcal{R}) & \text{iff } \mu(1-\mu)(\alpha_B - \alpha_A)^2 \pi^D > \frac{[c(\delta-\overline{\alpha})+\overline{\alpha}S]^2}{c(\delta-\overline{\alpha})}. \end{cases} \tag{3.18}$$

The intuition is as follows. If benefits from managerial decision-making are sufficiently low, i.e. $\delta \leq \alpha_1 < \overline{\alpha}$, the principal refrains from delegat-

[13]One example for a failure of organizational design due to missing complementarities are GM's investments in automation and flexibility in its production processes in the 1980s. As GM did not make complementary changes in other dimensions of internal organization, the investment eventually led to large losses (Roberts 2007, p.40).

ing any decision rights. Yet if there is no delegation the private benefits of a manager are zero and both promotion rules become equivalent. Thus, the principal is indifferent between the two. If $\delta \in (\alpha_1, \overline{\alpha}]$, the principal optimally chooses either partial or full delegation ($k^{\mathcal{P}} = \{\tilde{k}, 1\} > 0$) under a performance-based promotion and no delegation under a random promotion. But because $k^{\mathcal{P}} = \{\tilde{k}, 1\} > 0$, zero delegation is dominated under performance-based promotion and so is zero delegation under random promotion, i.e. $E\Pi^{\mathcal{P}}(\tilde{k}^{\mathcal{P}}) > E\Pi^{\mathcal{P}}(k^{\mathcal{P}} = 0) = E\Pi^{\mathcal{R}}(k^{\mathcal{R}} = 0$. If $\delta > \overline{\alpha}$, the principal fully delegates under a random promotion. Thus, if she also fully delegates under a performance-based promotion (i.e. when $\delta > \alpha_2$) the trade-off between motivation and selection determines the optimal promotion rule. A performance-based promotion is optimal if the selection effect is sufficiently weak. For partial delegation under a performance-based promotion (i.e. when $\overline{\alpha} \leq \delta < \alpha_2$), a random promotion gives higher payoffs from managerial decision-making, but lacks incentives and selection. Consequently, a sufficiently low selection effect implies that the gains from increasing worker incentives may outweigh lower managerial profits due to lower delegation.

3.4 Extensions

This section discusses the effects of monetary incentives for workers and managers on optimal delegation under performance-based promotions.[14]

3.4.1 Worker Wages

Additionally to their promotion prospects, workers receive wage payments dependent on the project outcome, namely w_S in case of success and w_F in case of failure. Workers remain risk-neutral. Furthermore, wages are constrained to be non-negative (e.g. because of workers' limited liability). Worker i's utility is then given by $u_i = p_i \cdot u_{m_i} + e_i \cdot w_S + (1 - e_i) \cdot w_F - \frac{c}{2} e_i^2 \geq 0$, and

[14] One may also wonder about a fully-integrated model with worker wages, manager wages and delegation. Yet Proposition 3.5 shows that manager wages are independent of delegation and consequently also of worker wages. Thus, results for a fully integrated model would not differ from the results presented in this section.

consequently optimal effort provision is given by $e_i^w = \frac{2(w_S - w_F) + u_{m_i}}{2c}$. The principal chooses jointly chooses $\{k, w_F, w_S\}$ to maximize expected profits of

$$E\Pi^w = \pi + k(\delta - \overline{\alpha})\pi^D + 2\overline{e}^w(S - w_S) + 2\mu(1-\mu)\frac{(p_B^w - p_A^w)(\Pi_B - \Pi_A)}{2} - w_F. \tag{3.19}$$

Introducing the bonus scheme has two effects. First, it motivates workers by giving additional incentives. But it also decreases potential profits from a worker's project success as the principal has to pay w_S to the worker in case of success. Both effects influence optimal delegation that is determined by Proposition 3.4.

Proposition 3.4.
The optimal organizational design with delegation and bonus pay, defined by $\{k^w, w_F^*, w_S^*\}$, *is given by*

$$(w_F^*, w_S^*) = (0, \frac{2S - k^w \cdot \overline{\alpha}\pi^D}{4}) \ \ \textit{and} \ \ k^w = \begin{cases} 1 & \textit{if } \delta \geq \alpha_2^w \\ \tilde{k}^w & \textit{if } \delta \in [\alpha_1^w, \alpha_2^w) \\ 0 & \textit{if } \delta < \alpha_1^w \end{cases} \tag{3.20}$$

with $\tilde{k}^w = \frac{c(\delta - \overline{\alpha}) + \frac{\overline{\alpha}S}{2}}{\mu(1-\mu)(\alpha_B - \alpha_A)^2\pi^D + \frac{\overline{\alpha}^2\pi^D}{4}}$, $\alpha_1^w = (1 - \frac{S}{2c})\overline{\alpha}$, *and* $\alpha_2^w = (1 - \frac{S}{2c})\overline{\alpha} + (\mu(1-\mu)(\alpha_B - \alpha_A)^2\pi^D + \frac{\overline{\alpha}^2\pi^D}{4})/c$.

There is a two-fold interaction between wages and delegation that is summarized by Corollaries 3.3 and 3.4 below. First, wages decrease profits in case of success which in turn lowers the optimal use of delegation. This is illustrated in Figure 3.3. Optimal delegation with additional wages (displayed by the dashed line) is weakly lower than optimal delegation without (straight line).

Corollary 3.3.
Introducing a bonus scheme lowers the use of delegation if $\delta \in (\alpha_1, \alpha_2^w)$.

Secondly, due to their joint use as worker incentive bonuses and delegation are substitutes as described by Corollary 3.4. For the intuition suppose there is a mean-preserving spread in the workforce heterogeneity, i.e. $\alpha_B - \alpha_A$

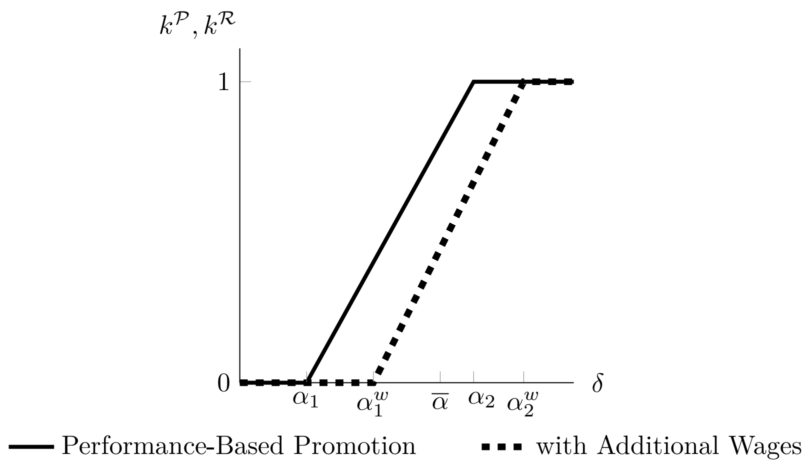


Figure 3.3: Optimal Delegation with and without Worker Wages

Note: This is an illustration of Proposition 3.4. For any given preference misalignment δ, optimal delegation under performance-based promotion with additional wage payments to the workers is weakly lower than without.

increases while $\overline{\alpha}$ remains constant. Then, the selection effect worsens and delegation becomes more costly. Consequently, the principal lowers delegation but increases bonuses to provide sufficient incentives to the workers.

Corollary 3.4.
Higher managerial discretion implies lower wages for subordinates, and vice versa.

3.4.2 Manager Wages

There is strong empirical evidence that "promotions are associated with large wage increases" (Waldman 2012, p.523). In contrast to private benefits from decision-making, the value of money is homogeneous for all workers. Thus, it prevents a negative selection effect. If that is the case, why not incentivize workers with wage increases instead of private benefits from decision-making? The answer is given below. Similar to a wage increase, delegation in this model is essentially a linear transfer of utility from principal to manager. The effects of both incentive devices are mainly similar. But delegation additionally comes with better decision-making by managers, thus increasing total surplus from management. Because costs from delegation, given by the selection effect, are negligible for low values of k, delegation becomes more profitable than similar increases in the manager's wage.

The principal offers an additional monetary prize of $\hat{w} \in [0, \hat{w}_{max}]$ to the promoted worker. A worker i's utility function is given by $\hat{u}_i = p_i(e_i, e_j) \cdot (\hat{w} + u_{m_i}) - \frac{c}{2}\hat{e}_i{}^2$. Note that the wage increase and private benefits become perfect substitutes in the worker's utility function. Given the resulting optimal effort level of $\hat{e}_i = \frac{w+u_{m_i}}{2c}$, the principal maximizes his expected profits of

$$E\hat{\Pi} = \pi + k(\delta - \overline{\alpha})\pi^D + \frac{(s\hat{w} + k\overline{\alpha}\pi^D)S}{c} + 2\mu(1-\mu)\frac{(p_B - p_A)(\Pi_B - \Pi_A)}{2} - \hat{w}. \tag{3.21}$$

Profit maximization over $\hat{w}$ and k gives the optimal organizational design with management wages. It is stated in Proposition 3.5.

Proposition 3.5.
The optimal organizational design with delegation and manager wages under performance-based promotions, defined by $\{\hat{w}^*, \hat{k}^*\}$, *is given by*

$$\{\hat{w}^*, \hat{k}^*\} = \begin{cases} \{0, 0\} & \text{if } S < c \cdot (1 - \frac{\delta}{\overline{\alpha}}) \\ \{0, k^{\mathcal{P}}\} & \text{if } c \cdot (1 - \frac{\delta}{\overline{\alpha}}) < S < c \\ \{\hat{w}_{max}, k^{\mathcal{P}}\} & \text{if } S > c \end{cases} \tag{3.22}$$

where $k^{\mathcal{P}}$ *is given in Proposition 3.2.*

There are three insights from Proposition 3.5. First, expected profits are linear in $\hat{w}$ and thus there exists a corner solution. Moreover there is no interdependence between manager wages and delegation. Thus, the optimal amount of delegation is the same as without manager wages and equivalent to optimal delegation in Proposition 3.2.

Furthermore, there exist an interesting relationship between manager wages and delegation on the extensive margin. If wages increase when switching from a working position to management (i.e. when $S > c$), the principal will also delegate. For the intuition, consider the marginal effects of wage increases and delegation on profits at zero. Remember from Section 3.3.1 that the selection effect is zero at $k = 0$. Thus, the marginal effect of delegation at zero is given by higher incentives and better managerial decision-making at the costs of loss of control: $\frac{\overline{\alpha}\pi^D \cdot S}{c} + \delta\pi^D - \overline{\alpha}\pi^D = \overline{\alpha}\pi^D(\frac{S}{c} - 1) + \delta\pi^D$. Compare this to the marginal effects from a wage increase on profits that stems from higher incentives and wage costs, i.e. $\frac{S}{c} - 1$.

Note the similarities between the two marginal effects. Delegation is essentially a utility transfer from principal to manager. By delegating a decision the principal gives, in expectation, $\overline{\alpha}\pi^D$ to the manager and thereby incentivizes workers. This translates into higher probabilities of success and the principal's expected profits are increased by $\frac{\overline{\alpha}\pi^D \cdot S}{c}$. This multiplier is equivalent for wage increases. Wages induce a marginal transfer of 1 from the principal to the worker, to receive higher expected profits of $\frac{S}{c}$. After accounting for the different "dimensions of utility" (monetary payment vs. private benefits from decision-making) the marginal effects are essentially the same. However delegation additionally increases total surplus as managers make better decisions, captured by $\delta\pi^D = \pi^D - \pi$. When delegating the principal keeps some share of that surplus increase. This gives additional incentives to delegate. Consequently, even when wage increases are not profitable, i.e. $\frac{S}{c} < 1$, delegation may still be. Proposition 3.5 therefore relies on the efficiency of delegation in this model. In contrast, if delegation decreases total surplus ($\pi^D < \pi$) the principal only delegates if gains from worker incentives are sufficiently large. Hence, the relationship between wage increases and delegation on the extensive margin would be reversed.

3.5 Conclusion

Many organizations rely on internal promotions to fill management positions, often based on employees' performance. Yet this wide-spread business practice can lead to suboptimal promotion decisions (Benson et al. 2018). Traditionally such findings were explained via differences in skill sets - employees who are good in worker tasks may not have the proper skills for management tasks. Here I pursue a different approach, showing that suboptimal promotions can optimally arise even without considering skill effects. This is due to the non-contractibility of management decision. It gives rise to managerial benefits that in turn affects workers behavior differently under performance-based promotions. Consequently, workers who make less profitable decisions as managers are promoted more likely.

I show that such interaction between managerial decision rights and worker behavior has various implications for organizational design. It affects how many decision rights should be delegated to a management position (Proposition 3.2), how promotions should be designed optimally (Corollary 3.2) and the joint decision of the two (Proposition 3.3). Moreover, optimal incentive schemes for both managers and workers can be linked to delegation, promotion and management selection (Propositions 3.4 and 3.5).

In the current model, workers want to become managers due to a non-contractibility of decision-making and subsequent private benefits. A recent literature has emphasized another reason for why individuals value decision rights. Fehr et al. (2013) and Bartling et al. (2014) find an *intrinsic* motivation for decision-making. In their experiments, individuals forgo money to make decisions themselves, without any instrumental or informational advantage. In an organizational context, these "power-hungry" individuals will influence for instance the optimal hierarchy of firms (Dessein and Holden 2019). Such preferences for power can also be put into the present context. Workers with higher intrinsic valuation of decision-rights have an higher incentive to work hard and thus will have a higher probability of getting promoted. Furthermore, as managers such individuals may try to hoard even more decision rights, for example by acquiring inefficiently many firms and becoming an empire-builder (Jensen 1986). In this case a trade-off between selection and incentives arises, similar to the current model. It will be interesting to further examine how the intrinsic valuation of decision rights interacts with different choices of organizational design.

APPENDICES

Appendix A

Appendix for Chapter 1

A.1 Proofs

Lemma 1.1

Lemma 1.1 follows from the maximization of the expected utilities at $t = 1$, see Equations (1.3) and (1.4). First, I derive the optimal effort levels given by Equation (1.5). The comparative statics stated in Lemma 1.1 are then straightforward. From $\max_e \quad Eu_{t1}^A = \left(1 - (1-\alpha)\pi\lambda^2\right) ev - \frac{c_e}{2}e^2 - \pi\lambda\alpha c_A$, it follows that $\frac{\partial Eu_{t1}^A}{\partial e} = (1-(1-\alpha)\pi\lambda^2)v - c_e e$ and the second derivative is negative. Therefore, $e^A = \frac{(1-(1-\alpha)\pi\lambda^2)v}{c_e}$. From $\max_e \quad Eu_{t1}^N = \left(1 - \pi\lambda^2\right) ev - \frac{c_e}{2}e^2$, it follows that $\frac{\partial Eu_{t1}^N}{\partial e} = (1-\pi\lambda^2)v - c_e e$ and the second derivative is negative. Therefore, $e^N = \frac{(1-\pi\lambda^2)v}{c_e}$.
The first statement of Lemma 1.1 follows from $e^A > e^N \; \forall \alpha > 0$. The second statement follows from $\frac{\partial e^A}{\partial \alpha} = \frac{\pi\lambda^2 v}{c_e} > 0$. □

Lemma 1.2

Recall $f(\lambda) := \left((2-(2-\alpha)\pi\lambda^2)\lambda v^2\right) - 2c_e c_A$ with $\lambda \in [0,1]$. First, I establish that f is concave in its domain $\lambda \in [0,1]$. The derivatives of $f(\lambda)$ with respect to λ are given by

$$\frac{\partial f(\lambda)}{\partial \lambda} = (2 - 3(2-\alpha)\pi\lambda^2)v^2 \quad \text{and} \quad \frac{\partial^2 f(\lambda)}{\partial \lambda^2} = -6(2-\alpha)\pi\lambda)v^2 \tag{A.1}$$

As $\frac{\partial^2 f(\lambda)}{\partial \lambda^2} < 0 \; \forall \lambda \in [0,1]$, it follows that $f(\lambda)$ is concave in λ with a unique global maximum at $\overline{\lambda} = \sqrt{\frac{2}{3(2-\alpha)\pi}}$. Therefore, $f(\lambda)$ is increasing for $\lambda < \overline{\lambda}$ and decreasing for $\lambda > \overline{\lambda}$. □

Lemma 1.3

First note that if $f(\overline{\lambda}) < 0$ no agent seeks advice. Lemma 1.2 states that $\overline{\lambda} = \sqrt{\frac{2}{3(2-\alpha)\pi}}$ and recall $f(\lambda) = \left((2-(2-\alpha)\pi\lambda^2)\lambda v^2\right) - 2c_e c_A$. Therefore, $f(\overline{\lambda})$ is given by

$$f(\overline{\lambda}) = \left(2 - (2-\alpha)\pi \cdot \frac{2}{3(2-\alpha)\pi}\right) \cdot \overline{\lambda}v^2 - 2c_e c_A = \frac{4}{3}\overline{\lambda}v^2 - 2c_e c_A \tag{A.2}$$

which is increasing in $\overline{\lambda}$. Therefore, if $f(\overline{\lambda} = 1) < 0$ it is also negative for all $\overline{\lambda} < 1$ and thus for all $\lambda \leq 1$. It follows that $f(\lambda) < 0$ if it holds that $\frac{4}{3}\overline{\lambda}v^2 - 2c_e c_A < 0$ with $\overline{\lambda} = 1$, i.e. if $c_A > \frac{2v^2}{3c_e}$. □

Lemma 1.4

Lemma 1.4 follows from comparing the optimal effort levels in Equation (1.17). I derive those by utility maximization. Expected utilities at $t = 1$ are denoted by

$$E\tilde{u}_{t1}^{A} = (1-\pi\lambda)(e(v+R^{N,S}) + (1-x)R^{N,F}) + \pi\lambda\left(e(v+R^{A,S}) + (1-x)R^{A,F} - c_A\right) - \frac{c_e}{2}e^2 \quad \text{(A.3)}$$

$$E\tilde{u}_{t1}^{N} = (1-\pi\lambda)(e(v+R^{N,S}) + (1-x)R^{N,F}) + \pi\lambda((1-\lambda)e(v+R^{N,S}) + \lambda R^{N,F}) - \frac{c_e}{2}e^2. \quad \text{(A.4)}$$

Simple re-formulation leads to the following, well-defined maximization problems:

$$\max_e \; E\tilde{u}_{t1}^{A} = e\cdot(v+d^N) + \pi\lambda\cdot x\left((d^A-d^N) + R^{A,F} - R^{N,F}\right) - \pi\lambda c_A - \frac{c_e}{2}e^2 \quad \text{(A.5)}$$

$$\max_e \; E\tilde{u}_{t1}^{N} = \left(1-\pi\lambda^2\right)\cdot e\cdot(v+d^N) - \frac{c_e}{2}e^2. \quad \text{(A.6)}$$

The first order derivatives are given by $\frac{\partial E\tilde{u}_{t1}^{A}}{\partial e} = (v+d^N) + \pi\lambda(d^A-d^N) - c_e e$ and $\frac{\partial E\tilde{u}_{t1}^{N}}{\partial e} = \left(1-\pi\lambda^2\right) - c_e e$; the second order derivatives are negative.
Thus $\tilde{e}^A = \frac{v+d^N+\pi\lambda(d^A-d^N)}{c_e}$ and $\tilde{e}^N = \frac{(1-\pi\lambda^2)(v+d^N)}{c_e}$.
The first statement of Lemma 1.4 follows from both effort levels increasing in d^A and d^N if $d^A, d^N > 0$. The second statement follows from $e^A - e^N = \frac{\pi\lambda}{c_e}(d^A - v - (1-\lambda)d^N)$. □

A.2 Equilibrium Analysis of Advice-Seeking with Reputation Concerns

There are four potential pure-strategy equilibria in the case of two types λ_l and λ_h. There are two pooling equilibria in which both type seek advice or do not seek advice. Also, there are two separating equilibrium in which only the high (low) type seeks advice and the low (high) type does not. The equilibrium concept is Perfect Bayesian Equilibrium (PBE). The D1 criterion (Cho and Kreps 1987) is applied as an equilibrium refinement to determine off-equilibrium beliefs.

Pooling on Advice-Seeking

In this equilibrium, both types seek advice. Therefore, type i exerts effort of

$$e_i^A = \frac{v + d^N + \pi\lambda_i(d^A - d^N)}{c_e}. \tag{A.7}$$

There are three possibilities regarding the ex-post reputation utility.[1]

1. Suppose $d^A = d^N$. In this case, both agents exert the same effort. Therefore, success or failure is no signal, and $\hat{\mu}^{a,S} = \hat{\mu}^{a,F}$ $\forall a$ and thus $d^A = d^N = 0$. As both agents would seek advice if possible and low types have a higher probability to need advice, this implies that $\hat{\mu}^{A,y} < \hat{\mu}^{N,y}$ $\forall y$. Then, there are three possibilities:

   - $\overline{\mu} > \hat{\mu}^{A,y}$: $R = 0$.
   - $\overline{\mu} \in [\hat{\mu}^{A,y}, \hat{\mu}^{N,y}]$: $R^N = r$.
   - $\overline{\mu} < \hat{\mu}^{A,y}$: $R^N = R^A = r$

   Two conditions then imply the pooling on advice-seeking equilibrium:

   (a) If $\overline{\mu} \in [\hat{\mu}^{A,y}, \hat{\mu}^{N,y}]$: $(2 - \pi\lambda^2)\lambda \cdot v^2 \geq 2c_e(c_A + r)$ $\forall\lambda \Leftrightarrow f(\lambda) \geq 2c_e r$ $\forall\lambda$.

   (b) If otherwise: $(2 - \pi\lambda^2)\lambda \cdot v^2 \geq 2c_e c_A$ $\forall\lambda \Leftrightarrow f(\lambda) \geq 0$ $\forall\lambda$.

   Therefore, condition for the equilibrium is:

$$(2 - \pi\lambda^2)\lambda \cdot v^2 \geq 2c_e(c_A + r) \text{ if } \overline{\mu} > \hat{\mu}^{A,y} = \frac{\pi\lambda_h}{\pi\lambda_h + (1-\pi)\lambda_l} \tag{A.8}$$

$$(2 - \pi\lambda^2)\lambda \cdot v^2 \geq 2c_e c_A \text{ if } \textit{otherwise}. \tag{A.9}$$

2. Suppose $d^A > d^N$. Then, the low type exerts higher effort (because he is more likely to receive advice and thus motivated more strongly to work hard) and success is a signal for being a low type. Therefore, $\hat{\mu}(a, S) < \hat{\mu}(a, F)$. As the low type is also more likely to need advice, it holds that $\hat{\mu}^{N,F}$ is the higher posterior belief and $\hat{\mu}^{A,S}$ is the lowest posterior belief. Therefore, the possibility in line with $d^A > d^N$ is that $d^A = 0$ and $d^N = -r$ and $\overline{\mu} \in (\max\{\hat{\mu}^{N,S}, \hat{\mu}^{A,F}\}, \hat{\mu}^{N,F}]$. In this case, the only conditions satisfying the pooling on advice-seeking equilibrium

   (a) $\overline{\mu} \in (\max\{\hat{\mu}^{N,S}, \hat{\mu}^{A,F}\}, \hat{\mu}^{N,F}]$

[1] The posterior beliefs in equilibrium are given by
$\hat{\mu}^{A,S} = \frac{\mu\lambda_h e_h^A}{\mu\lambda_h e_h^A + (1-\mu)\lambda_l e_l^A}$, $\hat{\mu}^{A,F} = \frac{\mu\lambda_h(1-e_h^A)}{\mu\lambda_h(1-e_h^A) + (1-\mu)\lambda_l(1-e_l^A)}$,
$\hat{\mu}^{N,S} = \frac{\mu(1-\pi\lambda_h)e_h^A}{\mu(1-\pi\lambda_h)e_h^A + (1-\mu)(1-\pi\lambda_l)e_l^A}$, and $\hat{\mu}^{N,F} = \frac{\mu(1-\pi\lambda_h)(1-e_h^A)}{\mu(1-\pi\lambda_h)(1-e_h^A) + (1-\mu)(1-\pi\lambda_l)(1-e_l^A)}$.

(b) $[(2-\pi\lambda^2)(v-r)+\pi\lambda\cdot r][\lambda v+(1-\lambda)r]+2c_e r \geq 2c_e c_A$
$\Leftrightarrow f(\lambda)+r(zr-wv)+2c_e r \geq 0$
where $w=2((2-\pi\lambda^2)\lambda-1)$ and $z=(1-\lambda)(2-\pi\lambda-\pi\lambda^2)$.

3. Suppose $d^A < d^N$. Then, the low type exerts higher effort and success is a signal for being a high type. Therefore, $\hat{\mu}(a,S) > \hat{\mu}(a,F)$. Together, this implies that $d^A=0, d^N=r$ and $\hat{\mu}^{N,S}$ must be the highest posterior belief and $\hat{\mu}^{A,F}$ must be the lowest posterior belief. In this case, the only conditions satisfying the pooling on advice-seeking equilibrium

   (a) $\overline{\mu} \in (\max\left\{\hat{\mu}^{A,S}, \hat{\mu}^{N,F}\right\}, \hat{\mu}^{N,S}]$

   (b) $[(2-\pi\lambda^2)(v+r)-\pi\lambda\cdot r][\lambda v-(1-\lambda)r] \geq 2c_e c_A \Leftrightarrow f(\lambda)-r(zr-wv) \geq 0$
   where $w=2((2-\pi\lambda^2)\lambda-1)$ and $z=(1-\lambda)(2-\pi\lambda-\pi\lambda^2)$.

Pooling on no Advice-Seeking

In this PBE, both types do not seek advice. Note that advice-seeking is off equilibrium path. The D1 criterion implies that $d^A=0$ (no updating on the off equilibrium path). The effort choice of type i are then given by

$$e_i^N \frac{(1-\pi\lambda_i^2)(v+d^N)}{c_e}. \tag{A.10}$$

Note that as $\lambda_h < \lambda_l$, it follows that $e_h^N > e_l^N$ and thus success is an indicator for the high type: $\hat{\mu}^{N,S} > \hat{\mu}^{N,F}$.[2] Therefore, there are two possibilities regarding ex-post reputation benefits.

- Suppose $\overline{\mu} < \hat{\mu}^{N,F}$ and thus $d^N=0$.
  - Suppose the high type has a higher incentive to deviate. Then, $R^{A,F}=r$ and the equilibrium condition is given by

$$(2-\pi\lambda_l^2)\lambda_l v < (2-\pi\lambda_h^2)\lambda_h v \leq 2c_e c_A \quad \Leftrightarrow \quad f(\lambda_l) \leq f(\lambda_h) \leq 0. \tag{A.11}$$

  - Suppose the low type has a higher incentive to deviate. Then, $R^{A,F}=0$ and the equilibrium condition is given by

$$\begin{aligned}&(2-\pi\lambda_h^2)\lambda_h v-2c_e r < (2-\pi\lambda_l^2)\lambda_l v-2c_e r \leq 2c_e c_A\\ &\Leftrightarrow \quad f(\lambda_l) \leq f(\lambda_h) \leq 2c_e r.\end{aligned} \tag{A.12}$$

- Suppose $d^N > 0$ and $\overline{\mu} \in (\hat{\mu}^{N,F}, \hat{\mu}^{N,S}]$. Then, $R^{N,S}=r$ and $R^{N,F}=0$.
  - Suppose the high type has a higher incentive to deviate. Then, $R^{A,F}=r$ and the equilibrium condition is given by

$$\begin{aligned}&[(2-\pi\lambda_l^2)(v+r)-\pi\lambda_l r][\lambda_l v-(1-\lambda_l)r] <\\ &< [(2-\pi\lambda_h^2)(v+r)-\pi\lambda_h r][\lambda_h v-(1-\lambda_h)r] \leq 2c_e c_A-2c_e r\\ &\Leftrightarrow \ f(\lambda_l)-r(zr+wv) \leq f(\lambda_h)-r(zr+wv) \leq -2cr.\end{aligned} \tag{A.13}$$

[2]The posterior beliefs in equilibrium given by $\hat{\mu}^{A,S} = \frac{\mu\lambda_h e_h^A}{\mu\lambda_h e_h^A+(1-\mu)\lambda_l e_A^l}$, $\hat{\mu}^{A,F} = \frac{\mu\lambda_h(1-e_h^A)}{\mu\lambda_h(1-e_h^A)+(1-\mu)\lambda_l(1-e_A^l)}$, $\hat{\mu}^{N,S} = \frac{\mu e_h^A}{\mu e_h^A+(1-\mu)e_A^l} = \frac{\mu(1-\pi\lambda_h^2)}{1-\pi[(1-\mu)\lambda_l^2+\mu\lambda_h^2]}$, and $\hat{\mu}^{N,F} = \frac{\mu(1-e_h^A)}{\mu(1-e_h^A)+(1-\mu)(1-e_A^l)} = \frac{\mu(1-e_h^A)}{(1-e_A^l)+\mu(e_l^A-e_h^A)}$.

– Suppose the low type has a higher incentive to deviate. Then, $R^{A,F}=0$ and the equilibrium condition is given by

$$\begin{aligned}&[(2-\pi\lambda_h^2)(v+r)-\pi\lambda_h r][\lambda_h v-(1-\lambda_h)r]<\\&<[(2-\pi\lambda_l^2)(v+r)-\pi\lambda_l r][\lambda_l v-(1-\lambda_l)r]\le 2c_e c_A\\&\Leftrightarrow\ f(\lambda_l)-r(zr+wv)\le f(\lambda_h)-r(zr+wv)\le 0.\end{aligned}\tag{A.14}$$

and again, $w=2((2-\pi\lambda^2)\lambda-1)$ and $z=(1-\lambda)(2-\pi\lambda-\pi\lambda^2)$.

Separation and the High Type Asks for Advice

Note in this PBE advice-seeking is fully informative: $\hat{\mu}^A=1$. However, no advice-seeking is not (with probability $1-\pi\lambda_h$ the high type does not need advice). This implies that $d^A=0$ and $R^{A,S}=R^{A,F}=r$. In this case, effort levels are given by

$$e_h^A=\frac{v+(1-\pi\lambda_h)d^N}{c_e}\quad\text{and}\quad e_l^N=\frac{(1-\pi\lambda_l^2)(v+d^N)}{c_e}.\tag{A.15}$$

Therefore, $e_h^A>e_l^N$ and success is a indicator for high ability. This implies $\hat{\mu}^{N,S}>\hat{\mu}^{N,F}$ and $d^N\ge 0$.[3]

- Suppose $\hat{\mu}\in(\hat{\mu}^{N,S},1]$ and $d^N=0$. In this case, the condition for the high type separation equilibrium is

$$\begin{aligned}&(2-\pi\lambda_h^2)\lambda_h v^2\ge 2c_e(c_A-r)>(2-\pi\lambda_l^2)\lambda_l v^2\\&\Leftrightarrow f(\lambda_h)\ge -2c_e r>f(\lambda_l)\end{aligned}\tag{A.16}$$

- Suppose $\hat{\mu}\in(\hat{\mu}^{N,F},\hat{\mu}^{N,S}]$ and $d^N=r$. In this case, the condition for the high type separation equilibrium is

$$\begin{aligned}&[(2-\pi\lambda_h^2)(v+r)-\pi\lambda_h r][\lambda_h v-(1-\lambda_h)r]\ge 2c_e(c_A-r)\\&\text{and}\ \ 2c_e(c_A-r)>[(2-\pi\lambda_l^2)(v+r)-\pi\lambda_l r][\lambda_l v-(1-\lambda_l)r]\\&\Leftrightarrow\ f(\lambda_h)-r(zr+wv)\ge -2c_e r>f(\lambda_l)-r(zr+wv)\end{aligned}\tag{A.17}$$

and again, $w=2((2-\pi\lambda^2)\lambda-1)$ and $z=(1-\lambda)(2-\pi\lambda-\pi\lambda^2)$.

Separation and the Low Type Asks for Advice

In this PBE, advice-seeking is fully informative for being a low type. Therefore, $\hat{\mu}^A=0$, $d^A=0$ and $R^{A,S}=R^{A,F}=0$. It follows the types provide following effort:

$$e_l^A=\frac{v+(1-\pi\lambda_l)d^N}{c_e}\quad\text{and}\quad e_h^N=\frac{(1-\pi\lambda_h^2)(v+d^N)}{c_e}.\tag{A.18}$$

This implies that the low type's effort level is higher if and only if $d^N<\frac{\lambda_h^2}{\lambda_l-\lambda_h^2}v$. In this case, $\hat{\mu}^{N,F}>\hat{\mu}^{N,S}$.[4]

[3] The posterior beliefs in equilibrium are given by $\hat{\mu}^{A,S}=\hat{\mu}^{A,F}=1$, $\hat{\mu}^{N,S}=\frac{\mu(1-\pi\lambda_h)e_h^A}{\mu(1-\pi\lambda_h)e_h^A+(1-\mu)e_A^l}$, and $\hat{\mu}^{N,F}=\frac{\mu(1-\pi\lambda_h)(1-e_h^A)}{\mu(1-\pi\lambda_h)(1-e_h^A)+(1-\mu)(1-e_A^l)}$.

[4] The posterior beliefs in equilibrium are given by $\hat{\mu}^{A,S}=\hat{\mu}^{A,F}=0$, $\hat{\mu}^{N,S}=\frac{\mu e_h^A}{\mu e_h^A+(1-\mu)(1-\pi\lambda_l)e_A^l}$, and $\hat{\mu}^{N,F}=\frac{\mu(1-e_h^A)}{\mu(1-e_h^A)+(1-\mu)(1-\pi\lambda_l)(1-e_A^l)}$.

- Suppose $\overline{\mu} \in (\hat{\mu}^{N,S}, \hat{\mu}^{N,F}]$ which implies $d^N < 0$. In this case, $R^{N,S} = 0$ and $R^{N,F} = r$ and the conditions for the low type separation equilibrium are

$$\begin{aligned} &[(2-\pi\lambda_l^2)(v-r)+\pi\lambda_l r][\lambda_l v+(1-\lambda_l)r]-2c_e r \geq 2c_e c_A \\ &\text{and } \ 2c_e c_A > [(2-\pi\lambda_h^2)(v-r)+\pi\lambda_h r][\lambda_h v+(1-\lambda_h)r]-2cr \\ &\Leftrightarrow \ f(\lambda_l)+r(zr+wv) \geq 2c_e c_A > f(\lambda_h)+r(zr+wv) \end{aligned} \tag{A.19}$$

- Suppose $\overline{\mu} < \hat{\mu}^{N,S}$ which implies $d^N = 0$ but $R^{N,S} = R^{N,F} = r$. There, and the condition for the low type separation equilibrium is

$$\begin{aligned} &(2-\pi\lambda_l^2)\lambda_l v^2 - 2c_e r \geq 2c_e c_A > (2-\pi\lambda_h^2)\lambda_h v^2 - 2c_e r \\ &\Leftrightarrow \ f(\lambda_l) \geq 2cr > f(\lambda_h) \end{aligned} \tag{A.20}$$

- Suppose $\overline{\mu} \in (\hat{\mu}^{N,F}, \hat{\mu}^{N,S}]$ which implies $d^N > 0$. This only holds if $\hat{\mu}^{N,F} < \hat{\mu}^{N,S}$ and thus $d^N = r > \frac{\lambda_h^2}{\lambda_l - \lambda_h^2} v$. In this case, $R^{N,S} = r$ and $R^{N,F} = 0$ and the conditions for the low type separation equilibrium are

$$\begin{aligned} &[(2-\pi\lambda_l^2)(v+r)-\pi\lambda_l r][\lambda_l v-(1-\lambda_l)r] \geq 2c_e c_A \\ &\text{and } \ 2c_e c_A > [(2-\pi\lambda_h^2)(v+r)-\pi\lambda_h r][\lambda_h v-(1-\lambda_h)r] \\ &\Leftrightarrow \ f(\lambda_l)-r(zr+wv) \geq 2cr > f(\lambda_h)-r(zr+wv) \end{aligned} \tag{A.21}$$

  and again, $w = 2((2-\pi\lambda^2)\lambda - 1)$ and $z = (1-\lambda)(2-\pi\lambda-\pi\lambda^2)$.

Appendix B

Appendix for Chapter 2

B.1 Additional Figures and Tables

Beliefs

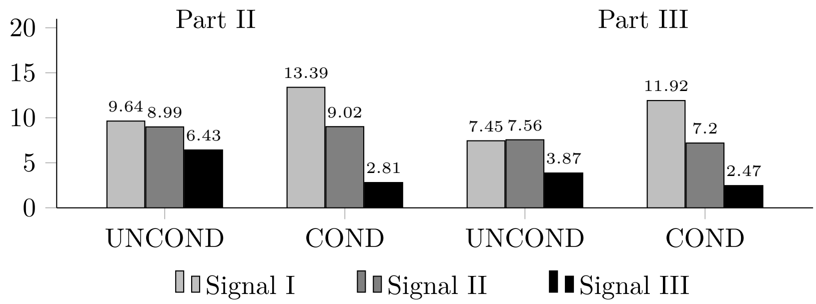
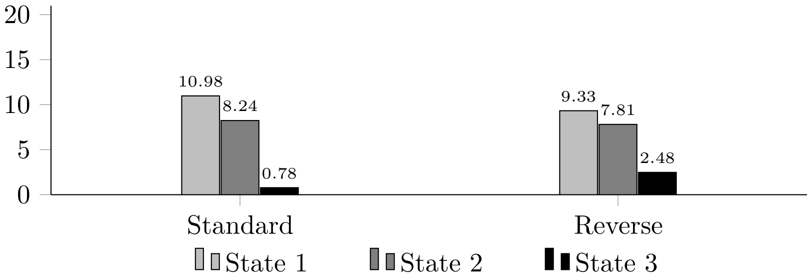


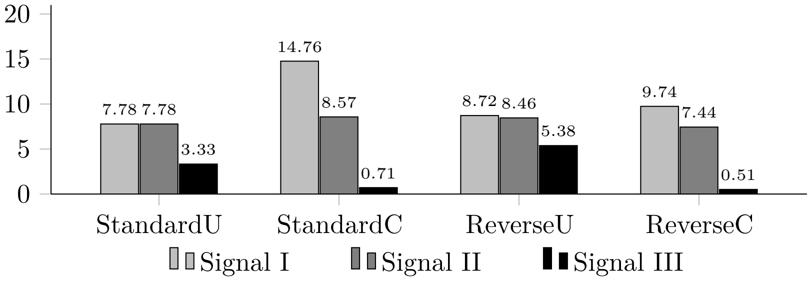
Figure B.1: Mean Beliefs for Each Signal in Part II and Part III

Order Effects

Figure B.2a depicts mean contributions in Part I, for each state conditional on the order. Standard denotes an order of "State 1, State 2, State 3". Reverse is "State 3, State 2, State 1". There is a positive effect of a state being introduced first. Contributions for State 3 are higher in Reverse than in Standard (p=0.0003). This could be because it may seem implausible to the subjects to play a public goods game without contributing. Consequently, they choose a nonzero contribution at the beginning. Contributions for State 1 are not significantly higher in Standard than Reverse (p=0.1927). Interestingly, the order effect is even more pronounced in Part II. Figure B.2b displays the mean contributions for each signal, conditional on the rule and the order. Rule have different effects, dependent on the order. More specifically, subjects seem to follow rules more strongly in their first contribution. If State 1 is presented first, contributions under the Conditional Rule for Signal I nearly double (p=0.0000). If it is presented last however, the Conditional Rule does not significantly change behavior (p=0.6645). Therefore, the order has a significant effect on Signal I contributions under the Conditional Rule (p=0.0175). Similar results holds for the Unconditional Rule and Signal III and could be explained by anchoring effects. Subjects are anchored by State 3 that it is inefficient to contribute and continue



(a) Mean Contributions for Each State in Part I, Conditional on Order



(b) Mean Contributions for Each Signal in Part II, Conditional on Order and Rule

Figure B.2: Mean Contributions, Conditional on the Order.

to do so throughout the experiment. Therefore, any rule has less effect on the subjects' contribution behavior.

Part III - Time Trend of Contributions

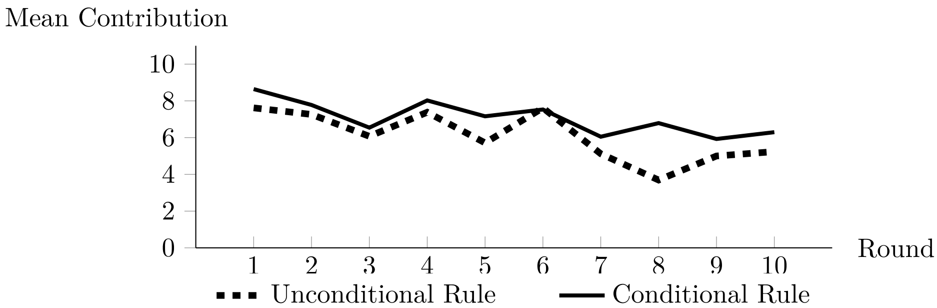


Figure B.3: Mean Contributions in Each Round in Part III

Part III - Additional Regression Analyses

Dependent variable: Indv. Contribution	Model 1	Model 2	Model 3	Model 4	Model 5
Round	-0.35***	-0.35***	-0.34***	-0.35***	-0.34***
	(0.670)	(0.076)	(0.077)	(0.067)	(0.068)
Conditional Rule		0.94	0.74		
		(0.744)	(0.774)		
Signal I	6.06***	8.70***	8.88***	6.14***	6.24***
	(0.803)	(0.812)	(0.791)	(0.836)	(0.820)
Signal II	6.14***	6.74***	6.95***	6.21***	6.33***
	(0.490)	(0.351)	(0.338)	(0.454)	(0.431)
Signal I * Cond. Rule	5.26***			3.81 **	3.67***
	(1.332)			(1.490)	(1.333)
Signal II * Cond. Rule	0.68			-0.26	-0.39
	(0.733)			(0.649)	(0.776)
Signal III * Cond. Rule				-1.26***	-1.60**
				(0.411)	(0.698)
Avr. Contr. Part I			0.37***		0.38***
			(0.059)		(0.589)
Constant	3.22***	2.67***	-2.45	3.80***	-1.09
	(0.526)	(0.655)	(1.520)	(0.453)	(1.498)
Controls			Yes		Yes
Number of obs.	1,620	1,620	1,590	1,620	1,590
Number of indv.	162	162	156	162	156
Number of group clusters	54	54	54	54	54

Table B.1: Further Estimation Models on Contributions in Part III

Note: Model 1: Fixed Effects GLS Model, Models 2-5: Pooled OLS Model. Robust standard errors in brackets, clustered at the group level. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.001$. Three groups were dropped because of a Rule B-group match or because of no match in the session with 18 subjects. Controls in Models 3 & 5 include age, gender, previous participations in experiments, self-reported math skills and measures for altruism, negative reciprocity, risk and trust. The question on gender was not answered by three participants, they were dropped in Models 3 & 5.

B.2 Instructions

In the following, I provide the original instructions as well as important exemplary computer screenshots for the experiment that was conducted in German. There were three printouts for the General Instructions, Part I and Part II. All other instructions were displayed on the computer screens. For exposition, screenshots are changed in size and format.

General Instructions (printed)

Instruktionen

Herzlichen Willkommen zu diesem Experiment und vielen Dank für Ihre Teilnahme. Bitte lesen Sie die folgenden Instruktionen sorgfältig durch.
Von nun an und für die gesamte Dauer des Experiments ist es Ihnen und allen anderen Teilnehmern untersagt, mit anderen Teilnehmern[1] zu kommunizieren, Ihr Mobiltelefon zu benutzen, oder andere Programme auf dem Computer zu starten. Falls Sie eine dieser Regeln brechen, müssen wir Sie leider vom Experiment und den Auszahlungen ausschließen.
Falls während des Experiments Fragen aufkommen, drücken Sie bitte den roten Knopf auf der Tastatur vor Ihnen. Einer der Experimentalleiter wird zu Ihnen kommen und Ihre Fragen privat beantworten.

Generelle Instruktionen und Bezahlung

Dieses Experiment befasst sich mit ökonomischem Entscheidungsverhalten. Es wird am Computer durchgeführt.
Das Experiment dauert ca. 75 Minuten. Es besteht aus drei Teilen. In jedem dieser Teile können Sie mit Ihren Entscheidungen Geld verdienen. Ihr Verdienst im Experiment kann sowohl von Ihren Entscheidungen als auch den Entscheidungen der anderen Experimentteilnehmer abhängen. Wie genau Ihr Verdienst von diesen Entscheidungen abhängt, wird Ihnen jeweils in den einzelnen Teilen erläutert.
Ihnen wird Ihr genauer Verdienst am Ende des Experiments mitgeteilt. Dazu werden Ihre einzelnen Verdienste aus jedem Teil des Experiments zusammengerechnet. Zusätzlich erhalten Sie 6 Euro für Ihre Teilnahme am Experiment. Ihr Gesamtverdienst wird Ihnen am Ende des Experiments privat und in bar ausgezahlt.
Während des Experiments werden wir Ihren Verdienst nicht in Euro bezeichnen, sondern in „Punkten“. Am Ende des Experiments wird Ihr Verdienst aus dem Experiment wieder in Euro umgerechnet. Dazu wird Ihr Gesamtverdienst durch 8 geteilt.
Die Umrechnungsrate beträgt also:

1 Punkt = 1/8 Euro

Ihr Gesamtverdienst in diesem Experiment in Euro lautet daher:

6 Euro + (Punkte aus dem Experiment / 8) Euro

Ihr Gesamtverdienst im Experiment wird bei der Auszahlung auf 50cent – Beträge aufgerundet.

Anonymität

[1] Zur besseren Verständlichkeit verwenden wir in den Instruktionen nur männliche Bezeichnungen. Diese sind geschlechtsneutral zu verstehen.

Alle Ihre Entscheidungen und Antworten bleiben anonym. Die anderen Teilnehmer erfahren nicht, welche Entscheidungen Sie getroffen haben und welcher Teilnehmer wie viel verdient hat.

Fragen

Wenn Sie gerade Fragen haben, heben Sie bitte die Hand. Haben Sie Fragen während des Experiments, können Sie die rote Taste auf der Tastatur vor Ihnen drücken. Einer der Experimentalleiter wird zu Ihnen kommen und Ihre Fragen privat beantworten.

Teil 1 – Instruktionen

Wir werden nun die Instruktionen für den ersten Teil des Experiments gemeinsam durchgehen.

Instructions for Part I (printed)

Teil 1 - Instruktionen

Zunächst werden Sie mit der grundlegenden Entscheidung vertraut gemacht, die Sie im Experiment treffen müssen. Bitte lesen Sie sich die Instruktionen gut durch.
Während des gesamten ersten Teils des Experiments sind Sie Teil einer Gruppe mit zwei weiteren Teilnehmern. Insgesamt besteht eine Gruppe also aus drei Mitgliedern. Alle Mitglieder und ihre Entscheidungen sind anonym.

Ihr Verdienst in Teil 1 des Experiments

In Teil 1 werden Sie und die anderen beiden Gruppenmitglieder Entscheidungen in **drei verschiedenen Situationen** treffen. Am Ende des Experiments wird eine der drei Situationen vom Computer zufällig (mit gleicher Wahrscheinlichkeit) ausgewählt. Ihr Verdienst in Teil 1 des Experiments hängt von Ihrer Entscheidung und den Entscheidungen der anderen beiden Gruppenmitglieder in dieser Situation ab. Wir erläutern nun zunächst die generelle Entscheidung in jeder Situation, bevor wir auf die einzelnen Situationen eingehen.

Ihre Einzahlungsentscheidung

Sie haben 20 Punkte zur Verfügung. Sie entscheiden darüber, wie viele dieser Punkte Sie in ein Gruppenkonto, und wie viele Punkte Sie in ein Privatkonto einzahlen. Jeder Punkt, den Sie nicht in das Gruppenkonto einzahlen, geht automatisch in das Privatkonto. Sie können in 10er Schritten einzahlen, also entweder 0 Punkte, 10 Punkte oder 20 Punkte.

Das Gruppenkonto

Sie können Punkte in das Gruppenkonto einzahlen. Daraufhin erhalten Sie eine Auszahlung aus dem Gruppenkonto. Wie hoch diese Auszahlung ist, hängt von der jeweiligen Situation ab und wird Ihnen später für jede Situation separat erläutert. Auch die anderen Gruppenmitglieder profitieren von Ihrer Einzahlung in das Gruppenkonto: sie erhalten eine ebenso hohe Auszahlung wie Sie. Genauso erhalten Sie eine Auszahlung für jede Einzahlung eines anderen Gruppenmitglieds.
Ihr Verdienst entsteht aus den Auszahlungen aus dem Gruppenkonto, die wiederum aus Ihrer Einzahlung und den Einzahlungen der anderen Gruppenmitglieder resultieren. Daher gilt:

Ihr Verdienst aus dem Gruppenkonto = Auszahlungen aus dem Gruppenkonto

Das Privatkonto Jeder Punkt, den Sie nicht in das Gruppenkonto einzahlen, wird in ein Privatkonto eingezahlt. Für jeden Punkt in Ihrem Privatkonto verdienen Sie einen Punkt. Wenn Sie beispielsweise 10 Punkte in das Privatkonto einzahlen, ist Ihr Verdienst aus dem Privatkonto 10 Punkte. Niemand außer Ihnen profitiert von einer Einzahlung in das Privatkonto. Daher gilt:

Ihr Verdienst aus dem Privatkonto = 20 Punkte - Ihre Einzahlung in das Gruppenkonto

Ihr Gesamtverdienst in einer Situation

Ihr Verdienst in einer Situation ist die Summe der Verdienste aus Gruppenkonto und Privatkonto:

Ihr Verdienst = 20 Punkte - Ihre Einzahlung in das Gruppenkonto + Auszahlungen aus dem Gruppenkonto

Fragen

Wenn Sie Fragen haben, können Sie nun die Hand heben. Haben Sie Fragen während des Experiments, drücken Sie bitte die rote Taste auf der Tastatur vor Ihnen. Ein Experimentalleiter wird zu Ihnen kommen und Ihre Frage privat beantworten.

Situation 1

Wir beginnen nun mit Situation 1 und Ihrer Einzahlungsentscheidung für diese Situation. Alle weiteren Instruktionen zu Teil 1 finden Sie auf dem Computerbildschirm.
Bitte geben Sie nun Ihr „Teilnehmerlabel“ auf dem Computerbildschirm ein.
Ihr Teilnehmerlabel ist gleich der Computer-Nummer, die groß an der Seitenwand an Ihrem Platz hängt.

Exemplary Screenshots for Part I

Situation 1

In Situation 1 lauten die jeweiligen Auszahlungen aus dem Gruppenkonto:

- Zahlt ein Mitglied **0 Punkte** in das Gruppenkonto ein, erhält jedes Mitglied eine Auszahlung von: **0 Punkte**.
- Zahlt ein Mitglied **10 Punkte** in das Gruppenkonto ein, erhält jedes Mitglied eine Auszahlung von: **5 Punkte**.
- Zahlt ein Mitglied **20 Punkte** in das Gruppenkonto ein, erhält jedes Mitglied eine Auszahlung von: **12 Punkte**.

Alle Zahlungen sind noch einmal in der Tabelle zusammengefasst:

Einzahlung eines Mitglieds in das Gruppenkonto	Auszahlung für jedes Mitglied in Situation 1
0 Punkte	0 Punkte
10 Punkte	5 Punkte
20 Punkte	12 Punkte

Beispiel 1

Zwei Gruppenmitglieder zahlen jeweils 0 Punkte, ein Gruppenmitglied zahlt 20 Punkte in das Gruppenkonto ein.
Für jede Einzahlung von 0 Punkten entsteht eine Auszahlung von 20 Punkten.
Für die Einzahlung von 20 Punkten entsteht eine Auszahlung von 12 Punkten.
Daher erhält jedes Gruppenmitglied (0 + 0 + 12) = 12 Punkte aus dem Gruppenkonto.
Dazu kommen noch die jeweiligen Verdienste aus dem Privatkonto.

Beispiel 2

Zwei Gruppenmitglieder zahlen jeweils 10 Punkte, ein Gruppenmitglied zahlt 20 Punkte in das Gruppenkonto ein.
Für jede Einzahlung von 10 Punkten entsteht eine Auszahlung von 5 Punkten.
Für die Einzahlung von 20 Punkten entsteht eine Auszahlung von 12 Punkten.
Daher erhält jedes Gruppenmitglied (5 + 5 + 12) = 22 Punkte aus dem Gruppenkonto.
Dazu kommen noch die jeweiligen Verdienste aus dem Privatkonto.

Verständnisfragen

Bitte klicken Sie auf "Weiter", um fortzufahren und Verständnisfragen zu Situation 1 zu beantworten.

Weiter

Description of State 1

Situation 1

Bitte treffen Sie nun Ihre Einzahlungsentscheidung in Situation 1.

Zusammenfassung der Situation

Einzahlung eines Mitglieds in das Gruppenkonto	Auszahlung für jedes Mitglied in Situation 1
0 Punkte	0 Punkte
10 Punkte	5 Punkte
20 Punkte	12 Punkte

Auszahlungstabelle - Situation 1

Ihr Verdienst ist immer gegeben durch:
20 Punkte - Ihre Einzahlung in das Gruppenkonto + Auszahlungen aus dem Gruppenkonto

Bitte treffen Sie nun Ihre Einzahlungsentscheidung in Situation 1.

Wie viele Punkte möchten Sie in das Gruppenkonto einzahlen?

- 0 Punkte
- 10 Punkte
- 20 Punkte

Weiter

Decision Screen - Contribution

Situation 1 - Einschätzung von Verhalten

Wie viele Punkte haben die anderen Gruppenmitglieder eingezahlt? Bitte geben Sie eine Einschätzung ab.

Zusammenfassung der Situation

Einzahlung eines Mitglieds in das Gruppenkonto	Auszahlung für jedes Mitglied in Situation 1
0 Punkte	0 Punkte
10 Punkte	5 Punkte
20 Punkte	12 Punkte

Auszahlungstabelle - Situation 1

Ihre Einschätzung

Bitte geben Sie Ihre Einschätzung über das Einzahlungsverhalten der anderen Gruppenmitglieder ab. Wenn Sie mit beiden Einschätzungen richtig liegen, erhöht sich Ihr Verdienst aus Situation 1 um zusätzliche 5 Punkte.

Ein Gruppenmitglied gibt...	Ein Gruppenmitglied gibt...
0 Punkte.	0 Punkte.
10 Punkte.	10 Punkte.
20 Punkte.	20 Punkte.

Ihre Einschätzung

Decision Screen - Beliefs

Instructions for Part II (printed)

Teil 2 – Instruktionen

Allgemeine Entscheidungssituation

In diesem Teil des Experiments befinden Sie sich in einer ähnlichen Entscheidungssituation wie in Teil 1: Während des gesamten zweiten Teils sind Sie Teil einer Gruppe mit zwei weiteren Teilnehmern. Wie im vorherigen Teil können Sie und die anderen beiden Gruppenmitglieder bis zu 20 Punkte in ein Gruppenkonto einzahlen.
Die jeweiligen Einzahlungen resultieren in Auszahlungen für jedes Gruppenmitglied. Auch Ihr Verdienst wird weiterhin bestimmt durch:

Ihr Verdienst = 20 Punkte - Ihre Einzahlung in das Gruppenkonto + Auszahlungen aus dem Gruppenkonto

Einmalige Einzahlung

Im Unterschied zu Teil 1 wird von jedem Gruppenmitglied nur einmalig in das Gruppenkonto eingezahlt. Eine Einzahlung resultiert dabei wieder in einer Auszahlung für jedes Gruppenmitglied. Die Höhe der Auszahlung hängt von der Situation ab, in der sich Ihre Gruppe befindet.
Es gibt drei mögliche Situationen. Das sind dieselben, die Sie in Teil 1 kennengelernt haben. Eine Übersicht über die drei Situationen finden Sie nochmals aufgelistet im Beiblatt zu diesen Instruktionen. Der Computer zieht zufällig (mit gleichen Wahrscheinlichkeiten) und **geheim**, welche der drei Situationen in Ihrer Gruppe relevant für die Auszahlungen aus dem Gruppenkonto sein wird.
Wenn Sie und Ihre Gruppenmitglieder ihre jeweiligen Einzahlungsentscheidungen treffen, wissen also weder Sie noch die anderen Gruppenmitglieder, welche Situation der Computer für Ihre Gruppe gezogen hat. Sie wissen demnach nicht mit Sicherheit, welche Auszahlung aus einer Einzahlung in das Gruppenkonto entsteht.
Am Ende des Experiments erfahren Sie sowohl welche Situation in diesem Teil in Ihrer Gruppe galt als auch wie viele Punkte die anderen beiden Gruppenmitglieder in das Gruppenkonto eingezahlt haben.

Signale

Auch wenn kein Gruppenmitglied weiß, in welcher Situation sich Ihre Gruppe befindet, erhalten sowohl Sie als auch die anderen Gruppenmitglieder jeweils vom Computer **"private"** Informationen über die vorherrschende Situation. Die Informationen sind privat, weil kein anderes Gruppenmitglied beobachten kann, welche Information Sie erhalten haben. Umgekehrt können auch Sie keine Informationen der anderen Gruppenmitglieder beobachten. Es gibt drei verschiedene Informationen. Wir nennen eine Information „Signal“. Es gibt Signal I, Signal II und Signal III.

Begriffsklärung
Das „korrekte Signal“ ist immer das Signal, das mit der Situation übereinstimmt. Ansonsten ist es ein „falsches Signal“.
In Situation 1 ist Signal I das korrekte Signal, Signal II und III sind falsche Signale. In Situation 2 ist Signal II das korrekte Signal, Signal I und III sind falsche Signale. In Situation 3 ist Signal III das korrekte Signal, Signal I und II sind falsche Signale.

Wir werden zunächst erst erklären, wie der Computer Signale generiert. Danach wird Ihnen erläutert, welche Rückschlüsse Sie aus einem Signal ziehen können.
Für jedes Gruppenmitglied simuliert der Computer **geheim, einzeln und unabhängig voneinander** den Wurf eines Würfels mit sechs Seiten. Der Computer wirft den Würfel also dreimal, einmal für jedes Gruppenmitglied. Bei jedem Wurf / Gruppenmitglied gilt:

- Zeigt der Würfel die Zahl 1 an, erhält das Gruppenmitglied ein falsches Signal. Welches der beiden falschen Signale dies ist, wird vom Computer zufällig bestimmt.
- Zeigt der Würfel eine der Zahlen 2, 3, 4, 5 oder 6 an, erhält das Gruppenmitglied das korrekte Signal.

Beispiele

1. Zuerst zieht der Computer Situation 2.
   Danach würfelt er nacheinander dreimal. Die folgenden Würfelzahlen entstehen:
   - Eine 1 für Gruppenmitglied A
   - Eine 5 für Gruppenmitglied B
   - Eine 3 für Gruppenmitglied C

   Das bedeutet: Gruppenmitglieder B und C erhalten das korrekte Signal II. Gruppenmitglied A erhält ein falsches Signal (I oder III). Der Computer bestimmt zufällig, dass A das Signal I erhält.
2. Zuerst zieht der Computer Situation 1.
   Danach würfelt er nacheinander dreimal. Die folgenden Würfelzahlen entstehen:
   - Eine 4 für Gruppenmitglied A
   - Eine 4 für Gruppenmitglied B
   - Eine 2 für Gruppenmitglied C

   Das bedeutet: Alle Gruppenmitglieder erhalten das korrekte Signal I.
3. Zuerst zieht der Computer Situation 3.
   Danach würfelt er nacheinander dreimal. Die folgenden Würfelzahlen entstehen:
   - Eine 1 für Gruppenmitglied A
   - Eine 6 für Gruppenmitglied B
   - Eine 1 für Gruppenmitglied C

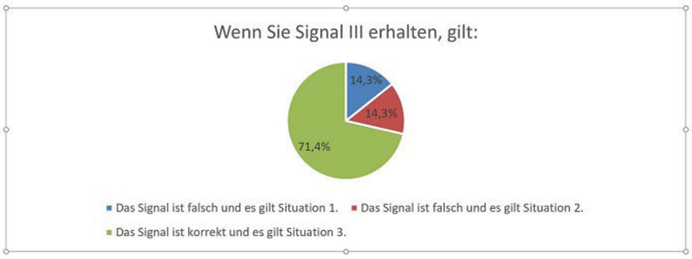
   Das bedeutet: Gruppenmitglieder A & C erhalten ein falsches Signal. Der Computer bestimmt zufällig und unabhängig für die beiden Gruppenmitglieder, dass A Signal I erhält und C Signal II erhält. Gruppenmitglied B erhält das korrekte Signal III.

Interpretation der Signale

Vor Ihrer Einzahlungsentscheidung erhalten Sie ein Signal. Da Sie weder die Situation oder die Würfelzahl des Computers beobachten können, wissen Sie nicht, ob dieses Signal das korrekte Signal oder ein falsches Signal ist. Mit den obigen Informationen kann man aber die folgenden Rückschlüsse ziehen:

1. Es ist am wahrscheinlichsten, dass das Signal korrekt ist. Dann wird auch die tatsächlich vorherrschende Situation in Ihrer Gruppe widergespiegelt.

2. Es ist am wahrscheinlichsten, dass Ihre Gruppenmitglieder dasselbe Signal wie Sie erhalten. Es ist jedoch auch möglich, dass die anderen Gruppenmitglieder andere Signale erhalten als Sie.
3. Außerdem kann man je nach Signal ausrechnen, mit welcher Wahrscheinlichkeit sich die Gruppe in einer Situation befindet. Wir haben dies graphisch für Sie aufbereitet. Die Kreisdiagramme zeigen dabei die Wahrscheinlichkeiten der verschiedenen Situationen an, wenn Sie ein Signal erhalten.



Zusammenfassung

- Der Computer zieht geheim eine Situation, die **in Ihrer Gruppe für alle Gruppenmitglieder** gilt. Daher ist der Zusammenhang zwischen Einzahlung und Auszahlung aus dem Gruppenkonto für alle Gruppenmitglieder derselbe.
- Dann erhalten die Gruppenmitglieder jeweils ein Signal. Dieses Signal kann korrekt oder falsch sein. Gruppenmitglieder können unterschiedliche Signale erhalten.
- Wenn Sie ein Signal erhalten, können Sie Rückschlüsse auf die in Ihrer Gruppe vorherrschende Situation ziehen. Dazu können Sie die Kreisdiagramme benutzen.
- Die Signale sind "privat": Kein Gruppenmitglied kann die Signale der anderen Gruppenmitglieder beobachten.

- Das bedeutet, Sie können nur begrenzt Rückschlüsse über das Verhalten der Gruppenmitglieder ziehen, wenn Sie am Ende des Experiments die Situation und die Einzahlungen der anderen beiden Gruppenmitglieder beobachtet haben. Denn Sie wissen nicht, welches Signal die anderen Gruppenmitglieder jeweils tatsächlich erhalten haben. Genauso können die anderen Gruppenmitglieder nur begrenzt Rückschlüsse über Ihr Verhalten ziehen.

Eine Gruppen-Regel für Ihre Gruppe

In Ihrer Gruppe gibt es außerdem eine Gruppen-Regel. Die Gruppen-Regel wird von einem anderen Teilnehmer im Experiment, „Teilnehmer R“, ausgewählt. Teilnehmer R ist in keiner Gruppe und trifft keine Einzahlungsentscheidung.
Teilnehmer R hat den Anreiz, eine Regel auszuwählen, die einen möglichst großen Gesamtverdienst in Ihrer Gruppe generiert. Der Gesamtverdienst bezeichnet die Summe der Verdienste aller Mitglieder einer Gruppe. Dies tut er, weil sein eigener Verdienst aus diesem Teil des Experiments unter anderem von diesem Gesamtverdienst abhängt. Es liegt in seinem Eigeninteresse, eine Regel auszuwählen, sodass der Gesamtverdienst so groß wie möglich ist. Teilnehmer R erhält außerdem weitere Informationen, die es ihm ermöglichen, eine gute Entscheidung zu treffen. Jedoch kennt auch Teilnehmer R die Situation in Ihrer Gruppe nicht.
Die Regel wird allen Gruppenmitgliedern eine Einzahlungsentscheidung für jedes Signal vorschreiben. **Die Regel ist jedoch nicht verpflichtend.** Das heißt, es gibt keine direkten negativen Konsequenzen für ein Gruppenmitglied, wenn sich dieses Gruppenmitglied nicht an die Regel hält.

Die nächsten Schritte und Ihr Verdienst in diesem Teil des Experiments

Zuerst wird Ihre Gruppen-Regel von einem Teilnehmer R ausgewählt und allen Mitgliedern in Ihrer Gruppe angezeigt.
Sie treffen eine Einzahlungsentscheidung für jedes der drei möglichen Signale. Jedoch wird nur eine dieser Entscheidungen tatsächlich ausgeführt:
Nachdem Sie die drei Einzahlungsentscheidungen getroffen haben, bestimmt der Computer die Situation, die in Ihrer Gruppe vorherrscht, und welches Signal jedes Gruppenmitglied erhält. Nur die Einzahlungsentscheidung für das tatsächlich erhaltene Signal wird ausgeführt.
Am Ende des Experiments erfahren Sie die Einzahlungen der anderen Gruppenmitglieder, welche Situation in Ihrer Gruppe gilt und Ihren resultierenden Verdienst in diesem Teil des Experiments. Dementsprechend erfahren auch die anderen Gruppenmitglieder die Höhe Ihrer Einzahlung.
Bitte warten Sie auf die Instruktionen des Experimentleiters, um fortzufahren. Sie werden dann zunächst Verständnisfragen beantworten.
Wenn Sie Fragen haben, heben Sie bitte jetzt Ihre Hand.

Exemplary Screenshots for Part II - Rulers

Ihre Rolle ist die von "Teilnehmer R"!

Ihre Rolle in diesem Teil des Experiments ist die von Teilnehmer R. Dementsprechend werden Sie in diesem Teil des Experiments keine Einzahlungsentscheidung vornehmen.

Stattdessen werden Sie zwei verschiedenen Gruppen zugewiesen. Ihr Aufgabe ist es in diesem Teil des Experiments ist es, für *jede* der Gruppen *jeweils eine* Gruppen-Regel auszuwählen. Ihr Ziel ist es, eine Regel auszuwählen, um einen möglichst großen Gruppenverdienst der jeweiligen Gruppe zu generieren. Der Gruppenverdienst bezeichnet die Summe der Verdienste aller Mitglieder einer Gruppe.

Ihr Verdienst

Ihr Verdienst aus diesem Teil des Experiments ist gleich dem durchschnittlichen Verdienst in einer der beiden Gruppen. Dazu wird der Gruppenverdienst einer Gruppe in diesem Teil, also die Summe der Verdienste aller Gruppenmitglieder, berechnet und dann durch Drei geteilt. Welche Gruppe releveant für Ihren Verdienst ist, wird zufällig bestimmt. Ihr Verdienst aus diesem Teil des Experiments lautet also:

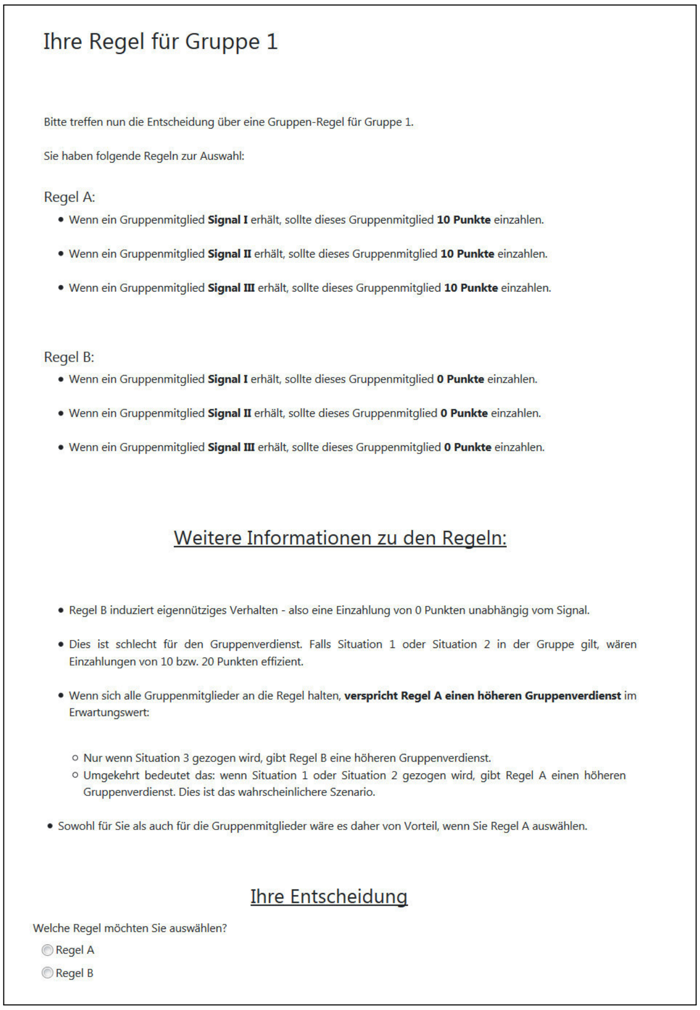
Ihr Verdienst = Gruppenverdienst einer der Ihnen zugewiesenen Gruppen / 3

Die Regel

Sie werden für jede Gruppe zwischen zwei verschiedenen Regeln auswählen können. Diese werden Ihnen, zusammen mit weiteren Informationen, die Regeln auf dem Auswahlbildschirm auf der nächsten Seite angezeigt.

Zunächst werden Sie eine Regel für die erste Gruppe, und dann für die zweite Gruppe auswählen. Bitte klicken Sie auf "Weiter", um eine Regel für die erste Gruppe auszuwählen.

Rulers' Additional Instructions



Ihre Regel für Gruppe 1

Bitte treffen nun die Entscheidung über eine Gruppen-Regel für Gruppe 1.

Sie haben folgende Regeln zur Auswahl:

Regel A:

- Wenn ein Gruppenmitglied **Signal I** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.
- Wenn ein Gruppenmitglied **Signal II** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.
- Wenn ein Gruppenmitglied **Signal III** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.

Regel B:

- Wenn ein Gruppenmitglied **Signal I** erhält, sollte dieses Gruppenmitglied **0 Punkte** einzahlen.
- Wenn ein Gruppenmitglied **Signal II** erhält, sollte dieses Gruppenmitglied **0 Punkte** einzahlen.
- Wenn ein Gruppenmitglied **Signal III** erhält, sollte dieses Gruppenmitglied **0 Punkte** einzahlen.

Weitere Informationen zu den Regeln:

- Regel B induziert eigennütziges Verhalten - also eine Einzahlung von 0 Punkten unabhängig vom Signal.
- Dies ist schlecht für den Gruppenverdienst. Falls Situation 1 oder Situation 2 in der Gruppe gilt, wären Einzahlungen von 10 bzw. 20 Punkten effizient.
- Wenn sich alle Gruppenmitglieder an die Regel halten, **verspricht Regel A einen höheren Gruppenverdienst** im Erwartungswert:
  - Nur wenn Situation 3 gezogen wird, gibt Regel B eine höheren Gruppenverdienst.
  - Umgekehrt bedeutet das: wenn Situation 1 oder Situation 2 gezogen wird, gibt Regel A einen höheren Gruppenverdienst. Dies ist das wahrscheinlichere Szenario.
- Sowohl für Sie als auch für die Gruppenmitglieder wäre es daher von Vorteil, wenn Sie Regel A auswählen.

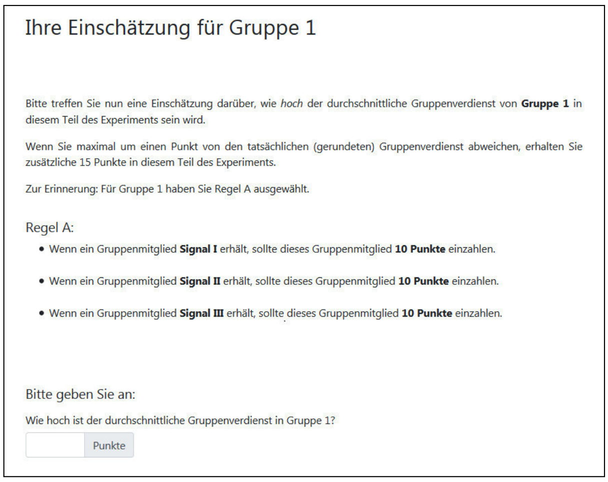
Ihre Entscheidung

Welche Regel möchten Sie auswählen?

○ Regel A

○ Regel B

Rulers' Rule Choice for Group 1



Ihre Einschätzung für Gruppe 1

Bitte treffen Sie nun eine Einschätzung darüber, wie *hoch* der durchschnittliche Gruppenverdienst von **Gruppe 1** in diesem Teil des Experiments sein wird.

Wenn Sie maximal um einen Punkt von den tatsächlichen (gerundeten) Gruppenverdienst abweichen, erhalten Sie zusätzliche 15 Punkte in diesem Teil des Experiments.

Zur Erinnerung: Für Gruppe 1 haben Sie Regel A ausgewählt.

Regel A:

- Wenn ein Gruppenmitglied **Signal I** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.
- Wenn ein Gruppenmitglied **Signal II** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.
- Wenn ein Gruppenmitglied **Signal III** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.

Bitte geben Sie an:

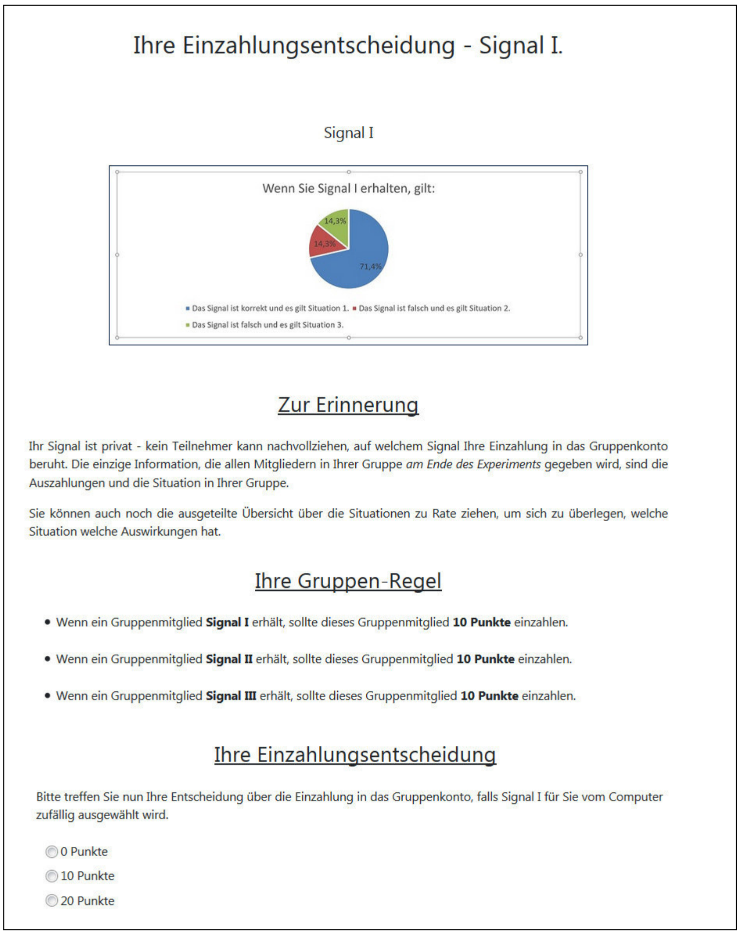
Wie hoch ist der durchschnittliche Gruppenverdienst in Gruppe 1?

[ ] Punkte

Elicitation of Rulers' Beliefs over Group Payoffs

Screenshots for Part II - Contributors

Display of Group Rule



Decision Screen - Contribution in Part II

Signal I - Einschätzung von Verhalten

Wie viele Punkte haben die anderen Gruppenmitglieder eingezahlt? Bitte geben Sie eine Einschätzung ab.

Wenn Sie mit beiden Einschätzungen richtig liegen, erhöht sich Ihr Verdienst um zusätzliche 5 Punkte, falls Signal I für Sie vom Computer zufällig ausgewählt wird.

Ein Gruppenmitglied gibt...	Ein Gruppenmitglied gibt...
◯ 0 Punkte. ◯ 10 Punkte. ◯ 20 Punkte.	◯ 0 Punkte. ◯ 10 Punkte. ◯ 20 Punkte.

Ihre Einschätzung

Decision Screen - Beliefs in Part II

Screenshots for Part III - Rulers

Teil 3 - Instruktionen.

In diesem Teil des Experiments sollen Sie verschiedene Einzahlungsentscheidungen eines "Individuum A" bewerten. Dazu sollen Sie bewerten, ob diese Entscheidung "sozial angemessen" oder "sozial unangemessen" ist.

Mit "sozial angemessen" ist Verhalten gemeint, **das die meisten Menschen als die "richtige" bzw. "ethische" Entscheidung bezeichnen würden**.

In jeder Ihrer Bewertungen bitten wir Sie darum, so ehrlich wie möglich zu antworten, basierend darauf, was Ihrer Meinung nach sozial angemessenes oder sozial unangemessenes Verhalten darstellt.

Beschreibung der generellen Einzahlungsentscheidung

"Individuum A" befindet sich in derselben Situation wie ein Gruppenmitglied im zweiten Teil dieses Experiments. Als Erinnerungshilfe können Sie sich die Instruktionen zu Teil 2 noch einmal durchlesen. Hier ist eine Zusammenfassung:

- "Individuum A" ist Teil einer Gruppe mit zwei anderen Teilnehmern. Jedes Gruppenmitglied kann bis zu 20 Punkte in ein Gruppenkonto einzahlen.
- Welche Auszahlung durch eine Einzahlung generiert wird, hängt von der jeweiligen Situation ab. Kein Gruppenmitglied kennt die Situation.
- "Individuum A" erhält (wie alle Gruppenmitglieder) eines der drei Signale als Hinweis über die vorherrschende Situation. Das Signal zeigt die Wahrscheinlichkeiten der jeweiligen Situation an, wie in den Instruktionen zu Teil 2 erläutert.
- Außerdem existiert eine Gruppen-Regel, die ein bestimmtes Verhalten aller Gruppenmitglieder vorschreibt. Diese Gruppen-Regel ist nicht verpflichtend.
- Daraufhin trifft "Individuum A" eine Einzahlungsentscheidung.

Ihre Informationen bei der Bewertung der Einzahlung

Um eine Bewertung der Entscheidungen von Individuum A vorzunehmen, erhalten Sie die folgenden Informationen:

- Die Einzahlungsentscheidung von Individuum A
- Die Gruppen-Regel der Gruppe von Individuum A
- Die tatsächliche Situation in der Gruppe
  - Merke: Individuum A kannte die Situation *nicht*, als es die Einzahlungsentscheidung traf. Individuum A erhielt nur eines der drei Signale. In 1 von 6 Fällen ist dies ein falsches Signal. In 5 von 6 Fällen ist dies das korrekte Signal.

Part III Instructions for Rulers (a)

Zur Illustration

Damit Sie ein Gefühl dafür bekommen, wie Ihre Aufgabe aussehen wird, haben wir Ihnen ein Bildschirm-Screenshot von Ihrer Bewertungsseite dargestellt. Wie Sie sehen, erhalten Sie zunächst Informationen über die Gruppen-Regel und die Situation der Gruppe.

Bitte bewerten Sie die folgende Einzahlungsentscheidung

Die Gruppen-Regel

Regel A

- Wenn ein Gruppenmitglied **Signal I** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.
- Wenn ein Gruppenmitglied **Signal II** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.
- Wenn ein Gruppenmitglied **Signal III** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.

Die Situation

In der Gruppe von Individuum A gilt: **Situation 1.**

Im Gegensatz zu Ihnen weiß Individuum A aber nicht, welche Situation in der Gruppe gilt. Individuum A erhält nur eine der drei Signale zur Information.

Sie wissen nicht, welches Signal "Individuum A" tatsächlich gesehen hat. Dadurch können Sie auch nicht mit Sicherheit sagen, auf welchem Signal die Einzahlungsentscheidung von Individuum A basiert.
Jedoch wissen Sie aus Teil 2, dass

- in 1 von 6 Fällen das Individuum ein falsches Signal sieht.
- in 5 von 6 Fällen das Individuum das korrekte Signal sieht.

Bitte schätzen Sie das folgende Verhalten ein:

Wenn Sie die folgenden Fragen beantworten, sollten Sie versuchen, so gut wie möglich die Antworten der anderen Gruppenmitglieder widerzuspiegeln!

Die Situation der Gruppe ist Situation 1 und "Individuum A" zahlt ... Punkte in das Gruppenkonto ein.	**Das ist sozial...**			
	sehr unangemessen	**unangemessen**	**angemessen**	**sehr angemessen**
0 Punkte	○	○	○	○
10 Punkte	○	○	○	○
20 Punkte	○	○	○	○

Weiter

In der Tabelle sehen Sie dann in der ersten Spalte die drei unterschiedlichen Einzahlungsbeträge, die "Individuum A" in das Gruppenkonto einzahlen kann. Für jeden der drei Einzahlungen sollen Sie angeben, ob Sie diese Einzahlung für

"sozial sehr unangemessen", "sozial unangemessen", "sozial angemessen" oder "sozial sehr angemessen"

halten.

Dazu sollten Sie sich auch bewusst sein, dass Individuum A sowohl ein korrektes als auch ein falsches Signal gesehen haben könnte.

Ihr Verdienst

Um Ihr Verdienst zu berechnen, werden Sie zufällig einem anderen Teilnehmer in diesem Teil des Experiments zugeteilt. Dieser Teilnehmer bewertet dieselben Einzahlungsentscheidungen wie Sie.

Nachdem alle Teilnehmer die Einzahlungsentscheidungen von Individuum A bewertet haben, wird der Computer einen der sechs Fälle und eine der potenziellen Einzahlungsentscheidungen zufällig auswählen. Dann wird Ihre Bewertung dieser Einzahlungsentscheidung mit der des Ihnen zugeteilten Teilnehmers verglichen.

Stimmt Ihre Bewertung mit der des anderen Teilnehmers überein, erhalten Sie weitere **20 Punkte** als Verdienst in diesem Teil des Experiments. Stimmt diese nicht überein, erhalten Sie **0 Punkte**.

Part III Instructions for Rulers (b) - it also includes a screenshot of the actual decision screen.

Screenshots for Part III - Contributors

Teil 3 - Instruktionen.

Allgemeine Entscheidungssituation

In diesem Teil des Experiments werden Sie eine Einzahlungsentscheidung wie in Teil 2 treffen. In Teil 3 wird diese jedoch wiederholt, für 10 Runden. Sie sind Teil einer Gruppe mit zwei anderen Teilnehmern. Diese Gruppe bleibt für die gesamten 10 Runden des dritten Teils bestehen.

Aufbau einer jeden Runde

In jeder Runde erhalten Sie und Ihre Gruppenmitgliedern jeweils ein individuelles, privates Signal. Wie zuvor werden die Signale durch unabhängige Würfelwürfe vom Computer bestimmt. Wie zuvor zeigt diese an, mit welcher Wahrscheinlichkeit eine Situation **für diese Runde** in Ihrer Gruppe gilt. Innerhalb einer Runde befinden sich alle Gruppenmitglieder in *derselben Situation*. Die Gruppenmitglieder können aber unterschiedliche Signale erhalten.

Sowohl die Situation in einer Gruppe als auch die individuellen Signale werden in jeder Runde neu bestimmt.

Sie können jederzeit auf die Instruktionen aus dem zweiten Teil zurückgreifen, die in diesem Teil in jeder Runde gelten.

Part III Instructions for Contributors (a)

Der strukturelle Aufbau einer jeden Runde

1. Der Computer zieht geheim und zufällig eine Situation, die in dieser Runde in Ihrer Gruppe gilt.
2. Dann erhält jedes Gruppenmitglied ein individuelles Signal.
3. Danach trifft jedes Gruppenmitglied die Einzahlungsentscheidung für diese Runde. Dabei erhalten alle Mitglieder zusätzlich eine Übersicht über den bisherigen Verlauf in Ihrer Gruppe. Diese Übersichtstabelle sieht aus wie folgt:

	Öffentliche Informationen				Private Informationen	
Runde	Situation	Einzahlungen der Gruppenmitglieder			Ihr Signal	Ihr Rundenverdienst
Runde x	Situation x	x Punkte	x Punkte	x Punkte	Signal x	x Punkte

x ist ein Platzhalter für die tatsächlichen Werte in der Tabelle.

Jedes Gruppenmitglied erhält dieselben öffentlichen Informationen über die Situation, die in den vorherigen Runde in der Gruppe galt, und die Einzahlungen aller Gruppenmitglieder. Außerdem erhält jedes Gruppenmitglied private Informationen über die eigenen vorherigen Signale und den eigenen vorherigen Rundenverdienst.

Das bedeutet im Umkehrschluss: kein Gruppenmitglied kann beobachten, welches Signal Sie gesehen haben und daher auf welcher Grundlage Sie Ihre Einzahlungsentscheidung getroffen haben. Jedoch erfährt jedes Gruppenmitglied am Ende einer Runde die Situation, die in dieser Runde in der Gruppe galt.

In den Instruktionen von Teil 2 ist beschrieben, dass mit einer Wahrscheinlichkeit von 1/6 (wenn eine 1 gewürfelt wird) ein falsches Signal generiert wird. Mit dieser Information sowohl Sie als auch die anderen Gruppenmitglieder sich überlegen, welches Signal jedes Gruppenmitglied gesehen hat, wenn in der Rundenzusammenfassung die Situation der Runde angezeigt wird.

Außerdem können die anderen Gruppenmitglieder zwar Ihre Einzahlung beobachten, Ihnen aber nicht persönlich zuordnen.

Die Regel

Auch in diesem Teil des Experiments gilt die Gruppen-Regel, die Teilnehmer R ausgewählt hat.

Diese Regel gilt für den gesamten dritten Teil des Experiments und wird Ihnen auf der nächsten Seite noch einmal angezeigt.

Ihr Verdienst

Am Ende des Experiments wird zufällig eine der zehn Runden ausgelost. Ihr Rundenverdienst aus dieser Runde ist dann Ihr Verdienst für den dritten Teil des Experiments.

Bitte klicken Sie nun auf "Weiter", um fortzufahren.

Part III Instructions for Contributors (b)

Runde 3 / 10

Ihr Signal

Zur Erinnerung: Kein anderes Gruppenmitglied kann beobachten, welches Signal Sie erhalten.

In dieser Runde erhalten Sie **Signal II** vom Computer.

Die Gruppen-Regel

In diesem Teil des Experiments lautet die Regel in Ihrer Gruppe:

- Wenn ein Gruppenmitglied **Signal I** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.
- Wenn ein Gruppenmitglied **Signal II** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.
- Wenn ein Gruppenmitglied **Signal III** erhält, sollte dieses Gruppenmitglied **10 Punkte** einzahlen.

Der bisherige Rundenverlauf

	Öffentliche Informationen				Private Informationen	
Runde	Situation	Einzahlungen der Gruppenmitglieder			Ihr Signal	Ihr Rundenverdienst
1	Situation 3	0 Punkte	20 Punkte	10 Punkte	Signal III	27 Punkte
2	Situation 1	20 Punkte	10 Punkte	10 Punkte	Signal III	22 Punkte

Ihre Entscheidung in dieser Runde

Wie viele Punkte möchten Sie in dieser Runde in das Gruppenkonto einzahlen?

- 0 Punkte
- 10 Punkte
- 20 Punkte

Decision Screen - Contribution in Part III

Instructions for Part IV

In Part IV, the following unincentivized survey questions (mostly building on Falk et al. (2018)) were asked on the computer screen.

1. Wie sehr sind Sie bereit oder nicht bereit, Risiken einzugehen? Bitte verwenden Sie die folgende Skala von 0 bis 10. Hier bedeutet 0, dass Sie 'überhaupt nicht bereit sind, Risiken einzugehen'. 10 bedeutet, dass Sie 'sehr bereit sind, Risiken einzugehen'. Sie können auch jede Zahl zwischen 0 und 10 verwenden, um anzugeben, wo sie sich auf der Skala sehen, in dem Sie (die Zahlen) 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, oder 10 verwenden. [0-10]
2. Stellen Sie sich die folgende Situation vor: Heute haben Sie unerwartet 1000 Euro erhalten. Wie viel von dem Geld würden Sie einem guten Zweck spenden? [0-1000]
3. Wir fragen Sie nun nach Ihrer Bereitschaft sich in einer bestimmten Art zu verhalten. Bitte verwenden Sie wieder eine Skala von 0 bis 10. 0 bedeutet 'überhaupt nicht bereit, dies zu tun' und 10 'sehr bereit, dies zu tun.' Sie können auch jede Zahl zwischen 0 und 10 verwenden, um anzugeben, wo sie sich auf der Skala sehen, in dem Sie (die Zahlen) 0, 1,2, 3, 4, 5, 6, 7, 8, 9, oder 10 verwenden.
   - Wie sehr wären Sie bereit, jemanden zu bestrafen, der Sie unfair behandelt, selbst wenn dies für Sie negative Konsequenzen haben würde? [0-10]
   - Wie sehr wären Sie bereit, jemanden zu bestrafen, der andere unfair behandelt, selbst wenn dies für Sie Kosten verursachen würde? [0-10]
   - Wie sehr wären Sie bereit, für einen guten Zweck zu geben, ohne etwas als Gegenleistung zu erwarten? [0-10]
4. Wie gut beschreibt jede der nachfolgenden Aussagen Sie als Person? Bitte verwenden Sie die Skala von 0 bis 10. 0 bedeutet 'beschreibt mich überhaupt nicht' und 10 'beschreibt mich perfekt'. Sie können auch jede Zahl zwischen 0 und 10 verwenden, um anzugeben, wo sie sich auf der Skala sehen, in dem Sie (die Zahlen) 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, oder 10 verwenden.
   - Wenn ich sehr ungerecht behandelt werde, räche ich mich bei der ersten Gelegenheit, selbst wenn Kosten entstehen, um das zu tun. [0-10]
   - Ich vermute, dass Leute nur die besten Absichten haben. [0-10]
   - Ich bin gut in Mathematik. [0-10]
5. Bitte geben Sie zuletzt noch folgende Informationen an.
   - An wie vielen Experimenten haben Sie bereits im MELESSA teilgenommen?
   - Bitte geben Sie Ihr Alter an. [0-99]
   - Bitte geben Sie Ihr Geschlecht an. [Männlich, Weiblich, Diverse, Keine Angabe]
   - Bitte wählen Sie Ihren höchsten Bildungsabschluss aus. [Hauptschulabschluss, Realschulabschluss, Abitur, B.Sc. / B.A., M.Sc. / M.A. / Diplom, Promotion, Anderes / Keine Angabe]
   - Falls Sie studieren, bitte geben Sie Ihr Studienfach an.

Preferences elicited by 1-4 were transformed into measured for risk, altruism, negative reciprocity and trust according to the procedure by Falk et al. (2018).

Appendix C

Appendix for Chapter 3

C.1 Optimal Promotion Tournaments

This appendix shortly discusses promotion rules that are a convex combination of random and performance-based promotions. For example, a firm could probabilistically switch its promotions practices between promoting the most successful worker or promoting based on other, performance-independent measures such as tenure.
For simplicity, fix k. Suppose the principal additionally chooses a probability $\rho \in [0,1]$ that determines the likelihood of a successful worker's promotion when his co-worker's project was a failure. $\rho = 1$ is equivalent to a (fully) performance-based promotion rule, $\rho = 0.5$ to random promotion. This gives a general promotion probability of $p_i^n = 0.5 + (\rho - 0.5)(e_i - \overline{e}_j)$ for worker i, resulting in optimal effort provision of $e_i^n = \frac{(\rho-0.5)u_{m_i}}{c}$. Note that for any $\rho \leq 0.5$, $e_i^n = 0$. Then, the principal's maximization problem is given by

$$\max_{\rho} E\Pi^n = \mu\Pi_B + (1-\mu)\Pi_A + 2\overline{e}^n(\rho) \cdot S + 2\mu(1-\mu)\frac{(p_B^n(\rho) - p_A^n(\rho))(\Pi_B - \Pi_A)}{2}. \tag{C.1}$$

Note that the effect of ρ only comes via workers' behavior. ρ increases average effort provision, but also increases the spread in promotion probabilities. However, due to the linearity of effort in ρ and the linearity of promotion probability in effort expected profits are linear in ρ. Thus, a "binary' solution arises. The principal either uses random promotion or performance-based promotion. A convex combination between the two is never optimal, stated in Proposition C.1. For the proof, see Appendix C.2.

Proposition C.1.
Either a random or a performance-based promotion is optimal, i.e. $\rho^n = \{0.5, 1\}$.

C.2 Proofs

Proposition 3.1

Note that expected profits (Equation (3.5)) are linear in k. They are increasing if $\delta > \overline{\alpha}$, constant if $\delta = \overline{\alpha}$ and decreasing if $\delta < \overline{\alpha}$. Proposition 3.1 follows immediately. □

Lemma 3.1

First note that expected utility is concave in e_i, and thus there is a unique maximum. The first order condition is given by $0.5 \cdot u_{m_i} - ce_i \overset{!}{=} 0$. Solving for the optimal effort level gives $e_i^{\mathcal{P}} = \frac{u_{m_i}}{2c} = k \cdot \frac{\overline{\alpha}_{m_i}\pi^D}{2c}$. □

Lemma 3.2

Under a heterogeneous workforce, as $p_i^h = 0.5 + 0.5(e_i - e_j)$, see Equation (3.8), we get

$$p_B^h - p_A^h = 0.5(e_B^{\mathcal{P}} - e_A^{\mathcal{P}}) - 0.5(e_A^{\mathcal{P}} - e_B^{\mathcal{P}}) = e_B^{\mathcal{P}} - e_A^{\mathcal{P}} = \frac{u_{m_B} - u_{m_A}}{2c} = k\frac{(\alpha_B - \alpha_A)\pi^D}{2c}. \tag{C.2}$$

Under a homogeneous workforce, $j = i$, and thus $p_i^{hom} = p_j^{hom} = 0.5$. □

Proposition 3.2

First, I show the simplification of the profit function to Equation (3.12), before maximizing Equation (3.12) over k.

$$\begin{aligned}
E\Pi =& (1-\mu)^2 (\Pi_A + 2e_A S) + 2\mu(1-\mu)\left(p_A^h \Pi_A + p_B^h \Pi_B + 2\overline{e}S\right) + \mu^2 (\Pi_B + 2e_B S) \\
=&\Pi_A \cdot \left((1-\mu)^2 + 2\mu(1-\mu)p_A^h\right) + \Pi_B \cdot \left(\mu^2 + 2\mu(1-\mu)p_B^h\right) \\
&+ 2S \cdot \left((1-\mu)^2 e_A + 2\mu(1-\mu)\overline{e} + \mu^2 e_B\right) \\
=&\Pi_A \cdot (1-\mu) + 2\mu(1-\mu)\frac{(e_A - e_B)}{2}\Pi_A + \Pi_B \cdot \mu + 2\mu(1-\mu)\frac{(e_B - e_A)}{2}\Pi_B \\
&+ 2S \cdot \left((1-\mu)^2 e_A + 2\mu(1-\mu)(\mu e_A + (1-\mu)e_B) + \mu^2 e_B\right) \\
=&(1-\mu)\Pi_A + \mu\Pi_B + 2S \cdot (\mu e_A + (1-\mu)e_B) + 2\mu(1-\mu)\frac{(e_B - e_A)}{2}(\Pi_B - \Pi_A) \\
=&(1-\mu)\Pi_A + \mu\Pi_B + 2\overline{e}S + 2\mu(1-\mu)\frac{(p_B^h - p_A^h)(\Pi_B - \Pi_A)}{2}.
\end{aligned} \tag{C.3}$$

Plugging in $(p_B - p_A)(\Pi_B - \Pi_A) = \frac{\left(k\overline{\alpha}\pi^D\right)^2}{2c}$ and $\overline{e} = \frac{k\overline{\alpha}\pi^D}{2c}$ gives Equation (3.12). The maximization problem is then

$$\max_k \; E\Pi^{\mathcal{P}} = \pi + k(\delta - \overline{\alpha})\pi^D + \frac{k\overline{\alpha}\pi^D}{c}S - \mu(1-\mu)\frac{\left(k(\alpha_B - \alpha_A)\pi^D\right)^2}{2c}. \tag{C.4}$$

The first- and second-order derivatives are given by

$$FD_k = (\delta - \overline{\alpha})\pi^D + \frac{\overline{\alpha}\pi^D}{c}S - k\mu(1-\mu)\left((\alpha_B - \alpha_A)\pi^D\right)^2/c \tag{C.5}$$

$$SD_k = -\mu(1-\mu)\left((\alpha_B - \alpha_A)\pi^D\right)^2/c < 0. \tag{C.6}$$

Thus, the profit function is concave in k. Note that due to the concavity of $E\Pi^{\mathcal{P}}$, if FD_k is negative at zero, $k = 0$ is optimal. Moreover, if FD_k is positive at 1, $k = 1$ is optimal. Furthermore, any $\tilde{k} \in [0, 1]$ with $FD_k(\tilde{k}) \overset{!}{=} 0$ is the unique interior solution to the maximization problem above.

1. FD_k is negative at zero if $(\delta - \overline{\alpha})\pi^D + \frac{\overline{\alpha}\pi^D}{c}S < 0$, or $\delta < \overline{\alpha} - \frac{S}{c}\overline{\alpha}$. Thus, $k^{\mathcal{P}} = 0$ if $\delta < \alpha_1 = (1 - \frac{S}{c})\overline{\alpha}$.

2. FD_k is non-negative at 1 if $(\delta - \overline{\alpha})\pi^D + \frac{\overline{\alpha}\pi^D}{c}S - \mu(1-\mu)\frac{\left((\alpha_B-\alpha_A)\pi^D\right)^2}{c} \geq 0$, or $\delta \geq (1 - \frac{S}{c})\overline{\alpha} + \mu(1-\mu)\frac{(\alpha_B-\alpha_A)^2\pi^D}{c} = \alpha_2$. Thus, $k^{\mathcal{P}} = 1$ if $\delta \geq \alpha_2$.

3. In any other case, we have an interior solution, implicitly given by $FD_k(\tilde{k}) \overset{!}{=} 0$ which gives $\tilde{k} = \frac{c(\delta-\overline{\alpha})+\overline{\alpha}S}{\mu(1-\mu)(\alpha_B-\alpha_A)^2\pi^D}$. □

Lemma 3.3

The selection effect is given by $-2\mu(1-\mu)\frac{\left(k(\alpha_B-\alpha_A)\pi^D\right)^2}{4c}$. First note that it is zero at $k = 0$ and negative for $k > 0$. Taking the first-order derivative w.r.t k gives $-\mu(1-\mu)k \cdot \frac{\left((\alpha_B-\alpha_A)\pi^D\right)^2}{c}$ which is negative for $k > 0$ and zero for $k = 0$. □

Corollary 3.1

Suppose $\alpha_2 \leq \overline{\alpha}$. Then, for any $\overline{\alpha} > \delta \geq \alpha_2$, $k^{\mathcal{P}} = 1 > 0 = k^{\mathcal{R}}$. This implies

$$\begin{aligned}
k^{\mathcal{P}} &= k^{\mathcal{R}} \text{ if } \delta \leq \alpha_1 \\
k^{\mathcal{P}} &> k^{\mathcal{R}} \text{ if } \delta \in (\alpha_1, \alpha_2) \\
k^{\mathcal{P}} &> k^{\mathcal{R}} \text{ if } \delta \in [\alpha_2, \overline{\alpha}) \\
k^{\mathcal{P}} &= k^{\mathcal{R}} \text{ if } \delta \geq \overline{\alpha}.
\end{aligned}$$

$\alpha_2 \leq \overline{\alpha}$ holds if $\mu(1-\mu)\left(\alpha_B - \alpha_A\right)^2\pi^D \leq \overline{\alpha}S < S$. □

Corollary 3.2

The difference in expected profits, given by Equation (3.14), is

$$\begin{aligned}
E\Pi^{\mathcal{P}} - E\Pi^{\mathcal{R}} =& 2\overline{e}S + 2\mu(1-\mu)(p_B^h - p_A^h)(\Pi_B - \Pi_A) \\
=& \frac{k\overline{\alpha}\pi^D S}{c} - \mu(1-\mu)\frac{\left(k(\alpha_B-\alpha_A)\pi^D\right)^2}{2c} \\
=& \frac{k\pi^D}{2c} \cdot (2\overline{\alpha}S - \mu(1-\mu)k\left(\alpha_B-\alpha_A\right)^2\pi^D),
\end{aligned} \tag{C.7}$$

which is positive whenever $k < \overline{k}^{\mathcal{P}} = \frac{2\overline{\alpha}S}{\mu(1-\mu)(\alpha_B-\alpha_A)^2\pi^D}$. □

Proposition 3.3

The proof for Proposition 3.3 is completed in several steps. Generally, we need to compare optimal expected profits under performance-based promotion against optimal expected profits under random promotion. As in both cases, the optimal degree of delegation is piecewise, we continue case by case. First note that $\alpha_1 < \min\{\overline{\alpha}, \alpha_2\}$.

1. Suppose $\delta \leq \alpha_1 < \min\{\overline{\alpha}, \alpha_2\}$.
   Then, $k^{\mathcal{P}} = k^{\mathcal{R}} = 0$ and thus $E\Pi^{\mathcal{P}} = E\Pi^{\mathcal{R}} = \pi$.

2. Suppose $\alpha_1 \leq \delta < \min\{\overline{\alpha}, \alpha_2\}$.
   Then, $k^{\mathcal{P}} = \tilde{k}$ and $k^{\mathcal{R}} = 0$ and thus the difference in expected payoffs is given by

$$\begin{aligned}
E\Pi^{\mathcal{P}}(\tilde{k}) - E\Pi^{\mathcal{R}}(0) &= \tilde{k} \cdot (\delta - \overline{\alpha})\pi^D + \tilde{k} \cdot \frac{\overline{\alpha}\pi^D}{c} S - \mu(1-\mu)\frac{\left(\tilde{k}(\alpha_B - \alpha_A)\pi^D\right)^2}{2c} \\
&= \tilde{k}\pi^D \left[\delta - (1 - \frac{S}{c})\overline{\alpha} - \mu(1-\mu)\tilde{k}\frac{(\alpha_B - \alpha_A)^2 \pi^D}{2}\right] \\
&= \tilde{k}\pi^D \left[\delta - (1 - \frac{S}{c})\overline{\alpha} - \mu(1-\mu)\frac{c(\delta - \overline{\alpha}) + \overline{\alpha}S}{\mu(1-\mu)\left(\alpha_B - \alpha_A\right)^2 \pi^D}\frac{(\alpha_B - \alpha_A)^2 \pi^D}{2}\right] \\
&= \tilde{k}\pi^D \left[\delta - (1 - \frac{S}{c})\overline{\alpha} - \frac{\delta - (1 - \frac{S}{c})\overline{\alpha}}{2}\right] \\
&= \tilde{k}\pi^D \left[\frac{\delta - (1 - \frac{S}{c})\overline{\alpha}}{2}\right] = \frac{\tilde{k}\pi^D}{2}\left[\delta - \alpha_1\right] > 0 \qquad \text{(C.8)}
\end{aligned}$$

3. Suppose $\alpha_2 < \overline{\alpha}$ and $\alpha_2 \leq \delta < \overline{\alpha}$.
   Then, $k^{\mathcal{P}} = 1$ and $k^{\mathcal{R}} = 0$ and thus the difference in expected payoffs is given by

$$\begin{aligned}
E\Pi^{\mathcal{P}}(1) - E\Pi^{\mathcal{R}}(0) &= (\delta - \overline{\alpha})\pi^D + \frac{\overline{\alpha}\pi^D}{c} S - \mu(1-\mu)\frac{\left((\alpha_B - \alpha_A)\pi^D\right)^2}{2c} \\
&= \pi^D \cdot \left[\delta - \left(\overline{\alpha}(1 - \frac{S}{c}) + \mu(1-\mu)\frac{(\alpha_B - \alpha_A)^2 \pi^D}{2c}\right)\right] \\
&= \pi^D \cdot \left[\delta - \alpha_2\right] > 0. \qquad \text{(C.9)}
\end{aligned}$$

4. Suppose $\alpha_2 > \overline{\alpha}$ and $\overline{\alpha} < \delta < \alpha_2$.
   Then, $k^{\mathcal{P}} = \tilde{k}$ and $k^{\mathcal{R}} = 1$ and thus

$$\begin{aligned}
E\Pi^{\mathcal{P}}(\tilde{k}) - E\Pi^{\mathcal{R}}(1) &= (\delta - \overline{\alpha})\pi^D(\tilde{k} - 1) + \tilde{k}\frac{\overline{\alpha}\pi^D}{c} S - \mu(1-\mu)\frac{\left(\tilde{k}(\alpha_B - \alpha_A)\pi^D\right)^2}{2c} \\
&= \frac{\pi^D}{c} \cdot \left[\frac{c(\delta - \overline{\alpha}) + \overline{\alpha}S}{2} \cdot \tilde{k} - (\delta - \overline{\alpha})c\right] \\
&= \frac{\pi^D}{c} \cdot \left[\frac{c(\delta - \overline{\alpha}) + \overline{\alpha}S}{2} \cdot \frac{c(\delta - \overline{\alpha}) + \overline{\alpha}S}{\mu(1-\mu)\left(\alpha_B - \alpha_A\right)^2 \pi^D} - (\delta - \overline{\alpha})c\right], \qquad \text{(C.10)}
\end{aligned}$$

   which is non-negative if and only if $\mu(1-\mu)\left(\alpha_B - \alpha_A\right)^2 \pi^D \leq \frac{[c(\delta - \overline{\alpha}) + \overline{\alpha}S]^2}{2c(\delta - \overline{\alpha})}$. Note that for $\alpha_2 > \overline{\alpha}$, it must hold that $\mu(1-\mu)\left(\alpha_B - \alpha_A\right)^2 \pi^D > \overline{\alpha}S$, but the two conditions are consistent.

5. $\delta \geq \max\{\overline{\alpha}, \alpha_2\}$.
   Then, $k^{\mathcal{P}} = k^{\mathcal{R}} = 1$ and thus

$$E\Pi^{\mathcal{P}}(1) - E\Pi^{\mathcal{R}}(1) = \frac{\overline{\alpha}\pi^D}{c} S - \mu(1-\mu)\frac{\left((\alpha_B - \alpha_A)\pi^D\right)^2}{2c}, \qquad \text{(C.11)}$$

   which is non-negative if and only if $\mu(1-\mu)\left(\alpha_B - \alpha_A\right)^2 \pi^D \leq 2\overline{\alpha}S$.

The first three cases can be summarized as:

$$\text{If } \delta \leq \alpha_1: \qquad (k^*, prom^*) = (0, \mathcal{P}) = (0, \mathcal{R}), \tag{C.12}$$
$$\text{If } \delta \in (\alpha_1, \overline{\alpha}]: \qquad (k^*, prom^*) = (k^{\mathcal{P}}, \mathcal{P}). \tag{C.13}$$

First note that $\overline{\alpha}S < 2\overline{\alpha}S < \frac{[c(\delta-\overline{\alpha})+\overline{\alpha}S]^2}{2c(\delta-\overline{\alpha})}$. Also for the two cases it holds that $\delta > \overline{\alpha}$. Further they can be summarized as follows.

1. $\mu(1-\mu)\left(\alpha_B - \alpha_A\right)^2 \pi^D < \overline{\alpha}S < 2\overline{\alpha}S$: $\alpha_2 < \overline{\alpha} < \delta$ and $(k^*, prom^*) = (1, \mathcal{P})$.
2. $\overline{\alpha}S \leq \mu(1-\mu)\left(\alpha_B - \alpha_A\right)^2 \pi^D \leq 2\overline{\alpha}S$: $\alpha_2 \geq \overline{\alpha}$ and $(k^*, prom^*) = (1, \mathcal{P})$.
3. $2\overline{\alpha}S < \mu(1-\mu)\left(\alpha_B - \alpha_A\right)^2 \pi^D < \frac{[c(\delta-\overline{\alpha})+\overline{\alpha}S]^2}{2c(\delta-\overline{\alpha})}$:
   $\alpha_2 \geq \overline{\alpha}$ and $(k^*, prom^*) = (\tilde{k}_1, \mathcal{P})$ where $\tilde{k}_1 = 1$ if $\delta \geq \alpha_2$ and $\tilde{k}$ otherwise.
4. $\mu(1-\mu)\left(\alpha_B - \alpha_A\right)^2 \pi^D > \frac{[c(\delta-\overline{\alpha})+\overline{\alpha}S]^2}{2c(\delta-\overline{\alpha})}$: $\alpha_2 \geq \overline{\alpha}$ and $(k^*, prom^*) = (1, \mathcal{R})$.

Proposition 3.3 is then a re-formulation of the above stated. □

Proposition 3.4

The profit maximization problem is given by

$$\max_{\{k, w_s, w_F\}} E\Pi^w = \pi + k(\delta - \overline{\alpha})\pi^D + 2\overline{e}^w(S - w_S) + \mu(1-\mu)(p_B^w - p_A^w)(\Pi_B - \Pi_A) - w_F, \tag{C.14}$$

which can be re-written (analogously as in Proposition 3.2) as

$$\begin{aligned}\max_{\{k, w_s, w_F\}} E\Pi^w =& \pi + k(\delta - \overline{\alpha})\pi^D + \frac{2(w_S - w_F) + k\overline{\alpha}\pi^D}{c}(S - w_S) \\ &- \mu(1-\mu)\frac{\left(k\left(\alpha_B - \alpha_A\right)\pi^D\right)^2}{2c} - w_F.\end{aligned} \tag{C.15}$$

Due to the workers' limited liability, $w_F = 0$ is optimal. Then, first- and second-order derivatives are then given by

$$FD_k = (\delta - \overline{\alpha})\pi^D + \frac{\overline{\alpha}\pi^D}{c}(S - w_S) - \mu(1-\mu)k\frac{\left((\alpha_B - \alpha_A)\pi^D\right)^2}{c} \tag{C.16}$$
$$FD_{w_S} = \frac{\left(2(S - w_S) - 2w_S + k\overline{\alpha}\pi^D\right)}{c} \tag{C.17}$$
$$SD_k = -\mu(1-\mu)\frac{\left((\alpha_B - \alpha_A)\pi^D\right)^2}{c} < 0 \tag{C.18}$$
$$SD_{w_S} = -2 < 0. \tag{C.19}$$

First, $w_S = \frac{2S - k\overline{\alpha}\pi^D}{4}$ is optimal, independent of k due to the independent concavity of profits in both parameters. Using this, we get

$$\begin{aligned}FD_k &= (\delta - \overline{\alpha})\pi^D + \frac{\overline{\alpha}\pi^D}{c}(S - \frac{2S - k\overline{\alpha}\pi^D}{4}) - \mu(1-\mu)k\frac{\left((\alpha_B - \alpha_A)\pi^D\right)^2}{c} \\ &= \delta - \overline{\alpha}(1 - \frac{S}{2c}) - k\left(\frac{\overline{\alpha}^2\pi^D}{4c} + \mu(1-\mu)\frac{\left(\alpha_B - \alpha_A\right)^2\pi^D}{c}\right)\end{aligned} \tag{C.20}$$

To find the optimal k I proceed as in the proof for Proposition 3.2, but the full procedure is omitted. Proposition 3.4 follows. $\square$

Corollary 3.3

We need to show that, on the extensive margin, the degree of delegation is lower with bonus schemes, thus for $k=0$, we have that $\alpha_{w1} > \alpha_1$ and for $k=1$, we have that $\alpha_{w2} > \alpha_2$. On the intensive margin, we need to show that $\tilde{k} > \tilde{k}^w$.

1. $\alpha_{w1} > \alpha_1$ holds as $\overline{\alpha} - \frac{\overline{\alpha}S}{2cx} > \overline{\alpha} - \frac{\overline{\alpha}S}{c}$.
2. $\alpha_{w2} > \alpha_2$ holds as $\alpha_{w2} = \alpha_2 + \frac{\overline{\alpha}S}{2c} + \frac{\left(\overline{\alpha}\pi^D\right)^2}{4c}$.
3. $\tilde{k} > \tilde{k}^w$ holds as

$$\tilde{k} = \frac{c(\delta-\overline{\alpha}) + \overline{\alpha}S}{\mu(1-\mu)\left(\alpha_B-\alpha_A\right)^2\pi^D} > \frac{\left(c(\delta-\overline{\alpha}) + \frac{\overline{\alpha}S}{2}\right)}{\mu(1-\mu)\left(\alpha_B-\alpha_A\right)^2\pi^D + \frac{\overline{\alpha}^2\pi^D}{4}} = \tilde{k}^w, \quad \text{(C.21)}$$

as the RHS's numerator is smaller and denominator is larger. $\square$

Corollary 3.4

This follows directly from Proposition 3.4 as $\frac{\partial w^S}{\partial k^w} = -\frac{\overline{\alpha}\pi^D}{4}$ for $\delta \in [\alpha_1^w, \alpha_2^w)$ (and otherwise k^w is constant). $\square$

Proposition 3.5

The first-order conditions with respect to $\hat{w}$ and k are given by

$$\frac{\partial E\hat{\Pi}}{\partial \hat{w}} = \frac{S}{c} - 1 \overset{!}{=} 0 \quad \text{(C.22)}$$

$$\frac{\partial E\hat{\Pi}}{\partial k} = (\delta-\overline{\alpha})\pi^D + \frac{\overline{\alpha}\pi^D}{c}S - k\mu(1-\mu)\frac{\left((\alpha_B-\alpha_A)\pi^D\right)^2}{c} \overset{!}{=} 0 \quad \text{(C.23)}$$

First note that the FOCs are independent. Secondly, $FOC_{\hat{k}}$ is the same as in Proposition 3.2 and the optimal amount of delegation is given by $k^{\mathcal{P}}$. Thirdly, to analyze when a positive amount of $\hat{w}$ or $\hat{k}$ is optimal, look at the behavior at $\hat{w}=0$, and $k=0$.

$$\frac{\partial E\hat{\Pi}}{\partial w}\,|_{\hat{w}=0} = \frac{S}{c} - 1 \quad \text{(C.24)}$$

$$\frac{\partial E\hat{\Pi}}{\partial k}\,|_{k=0} = \delta\pi^D + \overline{\alpha}\pi^D \cdot \left(\frac{S}{c} - 1\right). \quad \text{(C.25)}$$

Thus, $\frac{\partial E\hat{\Pi}}{\partial \hat{w}}\,|_{\hat{w}=0} > 0$ if and only if $S > c$ which in turn implies that $\frac{\partial E\hat{\Pi}}{\partial \hat{k}}\,|_{k=0} > 0$. Also, $\frac{\partial E\hat{\Pi}}{\partial \hat{k}}\,|_{k=0} > 0$ if and only if $S > c(1-\frac{\delta}{\overline{\alpha}})$. Taken together, the three cases stated in Proposition 3.5 arise. $\square$

Proposition C.1

First note $\overline{e}^n(\rho) = \frac{(\rho-0.5)k\overline{\alpha}\pi^D}{2c}$ and $p_B^n(\rho) - p_A^n(\rho) = \frac{(\rho-0.5)k(\alpha_B-\alpha_A)\pi^D}{2c}$. Then the first-order derivative of expected profits is $\frac{\partial E\Pi^n}{\partial \rho} = \frac{k\overline{\alpha}\pi^D S}{c} - \mu(1-\mu)\frac{\left(k(\alpha_B-\alpha_A)\pi^D\right)^2}{2c}$, which is independent of ρ. Thus a binary solution is optimal. Since $\rho < \frac{1}{2}$ would imply $e < 0$ we can restrict the possible set of solutions to $\rho^n \in \{\frac{1}{2}; 1\}$. Thus either random or performance-based promotions optimally emerge. $\square$

Bibliography

Acemoglu, **Daron**, **Georgy Egorov**, and **Konstantin Sonin** (2010). "Political Selection and Persistence of Bad Governments". *Quarterly Journal of Economics*, vol. 125 (4), pp. 1511–1575.

Aghion, **Philippe** and **Jean Tirole** (1997). "Formal and Real Authority in Organizations". *Journal of Political Economy*, vol. 105 (1), pp. 1–29.

Akerlof, **George A.** (1980). "A Theory of Social Custom, of which Unemployment may be One Consequence". *Quarterly Journal of Economics*, vol. 94 (4), pp. 749–775.

Alfaro, **Laura**, **Nick Bloom**, **Paola Conconi**, **Harald Fadinger**, **Patrick Legros**, **Andrew F. Newman**, **Raffaella Sadun**, and **John van Reenen** (2016). "All Together Now: Integration, Delegation and Management". Unpublished.

Altmann, **Steffen**, **Armin Falk**, **Simon Jäger**, and **Florian Zimmermann** (2018). "Learning about Job Search: A Field Experiment with Job Seekers in Germany". *Journal of Public Economics*, vol. 164, pp. 33–49.

Amabile, **Teresa M.**, **Colin M. Fisher**, and **Julianna Pillemer** (2014). "IDEO's Culture of Helping". *Harvard Business Review*, vol. 92 (1/2), pp. 54–61.

Amabile, **Teresa M.**, **Elizabeth A. Schatzel**, **Giovanni B. Moneta**, and **Steven J. Kramer** (2004). "Leader Behaviors and the Work Environment for Creativity: Perceived Leader Support". *Leadership Quarterly*, vol. 15 (1), pp. 5–32.

Andreoni, **James** and **B. Douglas Bernheim** (2009). "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects". *Econometrica*, vol. 77 (5), pp. 1607–1636.

Arrow, **Kenneth J.** (1964). "Control in Large Organizations". *Management Science*, vol. 10 (3), pp. 397–408.

— (1974). *The Limits of Organization.* New York City: WW Norton & Company.

Ashford, **Susan J.** and **L.L. Cummings** (1983). "Feedback as an Individual Resource: Personal Strategies of Creating Information". *Organizational Behavior and Human Performance*, vol. 32 (3), pp. 370–398.

Ashford, **Susan J.** and **Gregory B. Northcraft** (1992). "Conveying More (or Less) than We Realize: The Role of Impression-Management in Feedback-Seeking". *Organizational Behavior and Human Decision Processes*, vol. 53 (3), pp. 310–334.

Autor, **David**, **Frank Levy**, and **Richard J. Murnane** (2003). "The Skill Content of Recent Technological Change: An Empirical Exploration". *Quarterly Journal of Economics*, vol. 118 (4), pp. 1279–1333.

Autor, **David** and **Edward Price** (2013). "The Changing Task Composition of the US Labor Market: An Update of Autor, Levy, and Murnane (2003)". Unpublished, available at https://economics.mit.edu/files/9758.

Baker, **George**, **Michael Gibbs**, and **Bengt Holmstrom** (1994). "The Wage Policy of a Firm". *Quarterly Journal of Economics*, vol. 109 (4), pp. 921–955.

Baker, **George**, **Michael Jensen**, and **Kevin Murphy** (1988). "Compensation and Incentives: Practice vs. Theory". *Journal of Finance*, vol. 43 (3), pp. 593–616.

Bandiera, **Oriana**, **Iwan Barankay**, and **Imran Rasul** (2013). "Team Incentives: Evidence from a Firm Level Experiment". *Journal of the European Economic Association*, vol. 11 (5), pp. 1079–1114.

Barro, **Robert J.** (1973). "The Control of Politicians: An Economic Model". *Public Choice*, vol. 14, pp. 19–42.

Bartling, **Björn**, **Ernst Fehr**, and **Holger Herz** (2014). "The Intrinsic Value of Decision Rights". *Econometrica*, vol. 82 (6), pp. 2005–2039.

Bebchuk, **Lucian A.** and **Lars A. Stole** (1993). "Do Short-Term Objectives Lead to Under- or Overinvestment in Long-Term Projects?" *Journal of Finance*, vol. 48 (2), pp. 719–729.

Bénabou, **Roland**, **Armin Falk**, and **Jean Tirole** (2018). *Narratives, Imperatives, and Moral Reasoning*. Working Paper 24798. National Bureau of Economic Research.

Bénabou, **Roland** and **Jean Tirole** (2006). "Incentives and Prosocial Behavior". *American Economic Review*, vol. 96 (5), pp. 1652–1678.

Benson, **Alan**, **Danielle Li**, and **Kelly Shue** (2018). *Promotions and the Peter Principle*. Working Paper 24343. National Bureau of Economic Research.

Bernard, **Chester I.** (1938). *The Functions of the Executive.* Cambridge, MA: Harvard University Press.

Bernhardt, **Dan** (1995). "Strategic Promotion and Compensation". *Review of Economic Studies*, vol. 62 (2), pp. 315–339.

Bernheim, **B. Douglas** (1994). "A Theory of Conformity". *Journal of Political Economy*, vol. 102 (5), pp. 841–877.

Besley, **Timothy** (2005). "Political Selection". *Journal of Economic Perspectives*, vol. 19 (3), pp. 43–60.

Bester, **Helmut** and **Daniel Krähmer** (2008). "Delegation and Incentives". *RAND Journal of Economics*, vol. 39 (3), pp. 664–682.

Bohns, **Vanessa K.** and **Francis J. Flynn** (2010). ""Why Didn't You Just Ask?" Underestimating the Discomfort of Help-Seeking". *Journal of Experimental Social Psychology*, vol. 46 (2), pp. 402–409.

Bolino, **Mark C.** and **Adam M. Grant** (2016). "The Bright Side of Being Prosocial at Work, and the Dark Side, Too". *Academy of Management Annals*, vol. 10 (1), pp. 599–670.

Bolino, **Mark C.**, **Anthony C. Klotz**, **William H. Turnley**, and **Jaron Harvey** (2013). "Exploring the Dark Side of Organizational Citizenship Behavior". *Journal of Organizational Behavior*, vol. 34 (4), pp. 542–559.

Bonaccio, **Silvia** and **Reeshad S. Dalal** (2006). "Advice Taking and Decision-Making: An Integrative Literature Review, and Implications for the Organizational Sciences". *Organizational Behavior and Human Decision Processes*, vol. 101 (2), pp. 127–151.

Borgatti, **Stephen P.** and **Rob Cross** (2003). "A Relational View of Information Seeking and Learning in Social Networks". *Management Science*, vol. 49 (4), pp. 432–445.

Brandts, **Jordi**, **Valeska Groenert**, and **Christina Rott** (2015). "The Impact of Advice on Women's and Men's Selection into Competition". *Management Science*, vol. 61 (5), pp. 1018–1035.

Brin, **David** (1985). *The Postman.* New York City: Bantam Books.

Brooks, **Alison Wood**, **Francesca Gino**, and **Maurice E. Schweitzer** (2015). "Smart People Ask for (My) Advice: Seeking Advice Boosts Perceptions of Competence". *Management Science*, vol. 61 (6), pp. 1421–1435.

Brooks, **Alison Wood** and **Leslie K. John** (2018). "The Surprising Power of Questions." *Harvard Business Review*, vol. 96 (3), pp. 60–67.

Bursztyn, **Leonardo** and **Robert Jensen** (2017). "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure". *Annual Review of Economics*, vol. 9 (1), pp. 131–153.

Carnegie, **Dale** (1948). *How to Stop Worrying and Start Living*. New York City: Simon and Schuster.

Çelen, **Boğaçhan**, **Shachar Kariv**, and **Andrew Schotter** (2010). "An Experimental Test of Advice and Social Learning". *Management Science*, vol. 56 (10), pp. 1687–1701.

Chandrasekhar, **Arun G.**, **Benjamin Golub**, and **He Yang** (2018). *Signaling, Shame, and Silence in Social Learning*. Working Paper 25169. National Bureau of Economic Research.

Chaudhuri, **Ananish** (2011). "Sustaining Cooperation in Laboratory Public Goods Experiments: a Selective Survey of the Literature". *Experimental Economics*, vol. 14 (1), pp. 47–83.

Chaudhuri, **Ananish**, **Andrew Schotter**, and **Barry Sopher** (2008). "Talking Ourselves to Efficiency: Coordination in Inter-Generational Minimum Effort Games with Private, Almost Common and Common Knowledge of Advice". *Economic Journal*, vol. 119 (534), pp. 91–122.

Chen, **Daniel L.**, **Martin Schonger**, and **Chris Wickens** (2016). "oTree - An Open-Source Platform for Laboratory, Online, and Field Experiments". *Journal of Behavioral and Experimental Finance*, vol. 9, pp. 88–97.

Cho, **In-Koo** and **David M. Kreps** (1987). "Signaling Games and Stable Equilibria". *Quarterly Journal of Economics*, vol. 102 (2), pp. 179–221.

Cleavenger, **Dean**, **William L. Gardner**, and **Ketan Mhatre** (2007). "Help-Seeking: Testing the Effects of Task Interdependence and Normativeness on Employees' Propensity to Seek Help". *Journal of Business and Psychology*, vol. 21 (3), pp. 331–359.

Croson, **Rachel T. A.** and **Melanie Marks** (2001). "The Effect of Recommended Contributions in the Voluntary Provision of Public Goods". *Economic Inquiry*, vol. 39 (2), pp. 238–249.

Dana, **Jason**, **Roberto A. Weber**, and **Jason Xi Kuang** (2007). "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness". *Economic Theory*, vol. 33 (1), pp. 67–80.

Deming, **David J.** (2017). "The Growing Importance of Social Skills in the Labor Market". *Quarterly Journal of Economics*, vol. 132 (4), pp. 1593–1640.

Deming, **David J.** and **Lisa B. Kahn** (2018). "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals". *Journal of Labor Economics*, vol. 36 (S1), pp. 337–369.

DePaulo, **Bella M.** and **Jeffrey D. Fisher** (1980). "The Costs of Asking for Help". *Basic and Applied Social Psychology*, vol. 1 (1), pp. 23–35.

Dessein, **Wouter** (2002). "Authority and Communication in Organizations". *Review of Economic Studies*, vol. 69 (4), pp. 811–838.

Dessein, **Wouter** and **Richard Holden** (2019). *Organizations with Power-Hungry Agents*. Working Paper 13526. CEPR Discussion Paper.

DeVaro, **Jed**, **Antti Kauhanen**, and **Nelli Valmari** (2019). "Internal and External Hiring". *ILR Review*, vol. 72 (4), pp. 981–1008.

Duflo, **Esther** and **Emmanuel Saez** (2003). "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment". *Quarterly Journal of Economics*, vol. 118 (3), pp. 815–842.

Eckel, **Catherine C.** and **Philip J. Grossman** (2008). "Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices". *Journal of Economic Behavior & Organization*, vol. 68 (1), pp. 1–17.

Edmans, **Alex**, **Vivian W. Fang**, and **Katharina A. Lewellen** (2017). "Equity Vesting and Investment". *Review of Financial Studies*, vol. 30 (7), pp. 2229–2271.

Ellingsen, **Tore** and **Magnus Johannesson** (2008). "Pride and Prejudice: The Human Side of Incentive Theory". *American Economic Review*, vol. 98 (3), pp. 990–1008.

Ericson, **Keith Marzilli** and **David Laibson** (2019). "Intertemporal Choice". *Handbook of Behavioral Economics - Foundations and Applications, vol. 2*. Ed. by **B. Douglas Bernheim**, **Stefano DellaVigna**, and **David Laibson**. North-Holland, pp. 1–67.

Exley, **Christine L.** (2015). "Excusing Selfishness in Charitable Giving: The Role of Risk". *Review of Economic Studies*, vol. 83 (2), pp. 587–628.

Fairburn, **James A.** and **James M. Malcomson** (2001). "Performance, Promotion, and the Peter Principle". *Review of Economic Studies*, vol. 68 (1), pp. 45–66.

Falk, **Armin**, **Anke Becker**, **Thomas Dohmen**, **Benjamin Enke**, **David Huffman**, and **Uwe Sunde** (2018). "Global Evidence on Economic Preferences". *Quarterly Journal of Economics*, vol. 133 (4), pp. 1645–1692.

Fedyk, **Anastassia** (2018). “Asymmetric Naivete: Beliefs About Self-Control”. Working Paper, available at doi.org/10.2139/ssrn.2727499.

Fehr, **Ernst**, **Holger Herz**, and **Tom Wilkening** (2013). “The Lure of Authority: Motivation and Incentive Effects of Power”. *American Economic Review*, vol. 103 (4), pp. 1325–59.

Fischbacher, **Urs**, **Simon Gächter**, and **Ernst Fehr** (2001). “Are People Conditionally Cooperative? Evidence from a Public Goods Experiment”. *Economics Letters*, vol. 71 (3), pp. 397–404.

Fisher, **Colin M.**, **Julianna Pillemer**, and **Teresa M. Amabile** (2018). “Deep Help in Complex Project Work: Guiding and Path-Clearing Across Difficult Terrain”. *Academy of Management Journal*, vol. 61 (4), pp. 1524–1553.

Flynn, **Francis J.** and **Vanessa K. Bohns** (2008). “If You Need Help, Just Ask: Underestimating Compliance with Direct Requests for Help.” *Journal of Personality and Social Psychology*, vol. 95 (1), pp. 128–143.

Frederiksen, **Anders**, **Lisa B. Kahn**, and **Fabian Lange** (2019). *Supervisors and Performance Management Systems.* Working Paper 23351. National Bureau of Economic Research.

Frederiksen, **Anders**, **Fabian Lange**, and **Ben Kriechel** (2017). “Subjective Performance Evaluations and Employee Careers”. *Journal of Economic Behavior & Organization*, vol. 134, pp. 408–429.

Galbiati, **Roberto** and **Pietro Vertova** (2008). “Obligations and Cooperative Behaviour in Public Good Games”. *Games and Economic Behavior*, vol. 64 (1), pp. 146–170.

— (2014). “How Laws Affect Behavior: Obligations, Incentives and Cooperative Behavior”. *International Review of Law and Economics*, vol. 38, pp. 48–57.

Gardner, **Howard** (1983). *Frames of Mind: The Theory of Multiple Intelligences.* New York City: Basic Books.

Garicano, **Luis** (2000). “Hierarchies and the Organization of Knowledge in Production”. *Journal of Political Economy*, vol. 108 (5), pp. 874–904.

Gibbons, **Roberts** and **John Roberts** (2012). *The Handbook of Organizational Economics.* Princeton: Princeton University Press.

Goodall, **Amanda H.**, **Lawrence M. Kahn**, and **Andrew J. Oswald** (2011). “Why Do Leaders Matter? A Study of Expert Knowledge in a Superstar Setting”. *Journal of Economic Behavior & Organization*, vol. 77 (3), pp. 265–284.

Grabner, **Isabella** and **Frank Moers** (2013). "Managers' Choices of Performance Measures in Promotion Decisions: An Analysis of Alternative Job Assignments". *Journal of Accounting Research*, vol. 51 (5), pp. 1187–1220.

Grant, **Heidi** (2018). *Reinforcements: How to Get People to Help You.* Brighton: Harvard Business Review Press.

Greiner, **Ben** (2015). "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE". *Journal of the Economic Science Association*, vol. 1 (1), pp. 114–125.

Grodal, **Stine**, **Andrew J. Nelson**, and **Rosanne M. Siino** (2015). "Help-Seeking and Help-Giving as an Organizational Routine: Continual Engagement in Innovative Work". *Academy of Management Journal*, vol. 58 (1), pp. 136–168.

Grossman, **Zachary** (2014). "Strategic Ignorance and the Robustness of Social Preferences". *Management Science*, vol. 60 (11), pp. 2659–2665.

Grunewald, **Andreas**, **Emanuel Hansen**, and **Gert Pönitzsch** (2019). "Political Selection and the Optimal Concentration of Political Power". Working Paper.

Haas, **Martine R.** and **Morten T. Hansen** (2005). "When Using Knowledge Can Hurt Performance: The Value of Organizational Capabilities in a Management Consulting Company". *Strategic Management Journal*, vol. 26 (1), pp. 1–24.

Hansen, **Morten T.** (2009). *Collaboration: How Leaders Avoid the Traps, Build Common Ground, and Reap Big Results.* Brighton: Harvard Business Review Press.

Hargadon, **Andrew B.** and **Beth A. Bechky** (2006). "When Collections of Creatives Become Creative Collectives: A Field Study of Problem Solving at Work". *Organization Science*, vol. 17 (4), pp. 484–500.

Hayek, **Friedrich A.** (1944). *The Road to Serfdom.* Chicago: University of Chicago Press.

— (1945). "The Use of Knowledge in Society". *American Economic Review*, vol. 35 (4), pp. 519–530.

Headlam, **James W.** (1891). *Election by Lot at Athens.* Cambridge, UK: Cambridge University Press.

Hirshleifer, **Jack** (1989). "Conflict and Rent-Seeking Success Functions: Ratio vs. Difference Models of Relative Success". *Public Choice*, vol. 63 (2), pp. 101–112.

Hofmann, **David A.**, **Zhike Lei**, and **Adam M. Grant** (2009). "Seeking Help in the Shadow of Doubt: The Sensemaking Processes Underlying how Nurses Decide Whom to Ask for Advice". *Journal of Applied Psychology*, vol. 94 (5), pp. 1261–1274.

Hope, **Ole-Kristan** and **Wayne B. Thomas** (2008). "Managerial Empire Building and Firm Disclosure". *Journal of Accounting Research*, vol. 46 (3), pp. 591–626.

Ichniowski, **Casey** and **Kathryn Shaw** (2003). "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices". *Journal of Economic Perspectives*, vol. 17 (1), pp. 155–180.

Jensen, **Michael C.** (1986). "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers". *American Economic Review*, vol. 76 (2), pp. 323–329.

Jensen, **Robert** (2010). "The (Perceived) Returns to Education and the Demand for Schooling". *Quarterly Journal of Economics*, vol. 125 (2), pp. 515–548.

Johnson, **Zac D.** and **Sara LaBelle** (2017). "An Examination of Teacher Authenticity in the College Classroom". *Communication Education*, vol. 66 (4), pp. 423–439.

Ke, **Rongzhu**, **Jin Li**, and **Michael Powell** (2018). "Managing Careers in Organizations". *Journal of Labor Economics*, vol. 36 (1), pp. 197–252.

Knight, **Frank H.** (1938). "Lippmann's The Good Society". *Journal of Political Economy*, vol. 46 (6), pp. 864–872.

Koch, **Alexander K.** and **Julia Nafziger** (2012). "Job Assignments under Moral Hazard: The Peter Principle Revisited". *Journal of Economics & Management Strategy*, vol. 21 (4), pp. 1029–1059.

Krupka, **Erin L.** and **Roberto A. Weber** (2013). "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association*, vol. 11 (3), pp. 495–524.

Lazear, **Edward P.** (2004). "The Peter Principle. A Theory of Decline". *Journal of Political Economy*, vol. 112 (S1), S141–S163.

Lazear, **Edward P.**, **Kathryn L. Shaw**, and **Christopher T. Stanton** (2015). "The Value of Bosses". *Journal of Labor Economics*, vol. 33 (4), pp. 823–861.

Lee, **Fiona** (1997). "When the Going Gets Tough, Do the Tough Ask for Help? Help Seeking and Power Motivation in Organizations". *Organizational Behavior and Human Decision Processes*, vol. 72 (3), pp. 336–363.

— (2002). "The Social Costs of Seeking Help". *Journal of Applied Behavioral Science*, vol. 38 (1), pp. 17–35.

Levy, **Gilat** (2004). "Anti-Herding and Strategic Consultation". *European Economic Review*, vol. 48 (3), pp. 503–525.

Maskin, **Eric** and **Jean Tirole** (2004). "The Politician and the Judge: Accountability in Government". *American Economic Review*, vol. 94 (4), pp. 1034–1054.

Milgrom, **Paul** and **John Roberts** (1990). "The Economics of Modern Manufacturing: Technology, Strategy, and Organization". *American Economic Review*, vol. 80 (3), pp. 511–528.

— (1992). *Economics, Organization and Management*. Englewood Cliffs, N.J.: Prentice-Hall.

Morrison, **Elizabeth Wolfe** and **Robert J. Bies** (1991). "Impression Management in the Feedback-Seeking Process: A Literature Review and Research Agenda". *Academy of Management Review*, vol. 16 (3), pp. 522–541.

Mueller, **Jennifer S.** (2012). "Why Individuals in Larger Teams Perform Worse". *Organizational Behavior and Human Decision Processes*, vol. 117 (1), pp. 111–124.

Mueller, **Jennifer S.** and **Dishan Kamdar** (2011). "Why Seeking Help from Teammates is a Blessing and a Curse: A Theory of Help Seeking and Individual Creativity in Team Contexts." *Journal of Applied Psychology*, vol. 96 (2), pp. 263–276.

Newark, **Daniel A.**, **Vanessa K. Bohns**, and **Francis J. Flynn** (2017). "A Helping Hand is Hard at Work: Help-Seekers' Underestimation of Helpers' Effort". *Organizational Behavior and Human Decision Processes*, vol. 139, pp. 18–29.

Perlow, **Leslie A.**, **Constance Noonan Hadley**, and **Eunice Eun** (2017). "Stop the Meeting Madness: How to Free Up Time for Meaningful Work." *Harvard Business Review*, vol. 95 (4), pp. 62–69.

Persson, **Torsten**, **Gérard Roland**, and **Guido Tabellini** (1997). "Separation of Powers and Political Accountability". *Quarterly Journal of Economics*, vol. 112 (4), pp. 1163–1202.

Peter, **Lawrence J.** and **Raymond Hull** (1969). *The Peter Principle.* New York City: William Morrow & Company, Inc.

Podsakoff, **Nathan P.**, **Steven W. Whiting**, **Philip M. Podsakoff**, and **Brian D. Blume** (2009). "Individual- and Organizational-Level Consequences of Organizational Citizenship Behaviors: A Meta-Analysis." *Journal of Applied Psychology*, vol. 94 (1), pp. 122–141.

Prendergast, **Canice** (1999). "The Provision of Incentives in Firms". *Journal of Economic Literature*, vol. 37 (1), pp. 7–63.

Rajan, **Raghuram G.** and **Julie Wulf** (2006). "The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies". *Review of Economics and Statistics*, vol. 88 (4), pp. 759–773.

Roberts, **John** (2007). *The Modern Firm.* Oxford: Oxford University Press.

Rodrik, **Dani** (2015). *Economics Rules: The Rights and Wrongs of the Dismal Science.* New York City: WW Norton & Company.

Rogelberg, **Steven G.**, **Cliff Scott**, and **John Kello** (2007). "The Science and Fiction of Meetings." *MIT Sloan Management Review*, vol. 48 (2), pp. 18–21.

Rosen, **Sherwin** (1986). "Prizes and Incentives in Elimination Tournaments". *American Economic Review*, vol. 76 (4), pp. 701–715.

Ryan, **Allison M.**, **Paul R. Pintrich**, and **Carol Midgley** (2001). "Avoiding Seeking Help in the Classroom: Who and Why?" *Educational Psychology Review*, vol. 13 (2), pp. 93–114.

Schotter, **Andrew** (2003). "Decision Making with Naive Advice". *American Economic Review*, vol. 93 (2), pp. 196–201.

Schotter, **Andrew** and **Barry Sopher** (2006). "Trust and Trustworthiness in Games: An Experimental Study of Intergenerational Advice". *Experimental Economics*, vol. 9 (2), pp. 123–145.

Schöttner, **Anja** and **Veikko Thiele** (2010). "Promotion Tournaments and Individual Performance Pay". *Journal of Economics & Management Strategy*, vol. 19 (3), pp. 699–731.

Spence, **Michael** (1973). "Job Market Signaling". *Quarterly Journal of Economics*, vol. 87 (3), pp. 355–374.

Sunstein, **Cass R.** (1996). "On the Expressive Function of Law". *University of Pennsylvania Law Review*, vol. 144 (5), pp. 2021–2053.

Tessler, **Richard C.** and **Shalom H. Schwartz** (1972). “Help Seeking, Self-Esteem, and Achievement Motivation: An Attributional Analysis.” *Journal of Personality and Social Psychology*, vol. 21 (3), pp. 318–326.

Thompson, **Phillip S.** and **Mark C. Bolino** (2018). “Negative Beliefs about Accepting Coworker Help: Implications for Employee Attitudes, Job performance, and Reputation.” *Journal of Applied Psychology*, vol. 103 (8), pp. 842–866.

Tyran, **Jean-Robert** and **Lars P. Feld** (2006). “Achieving Compliance when Legal Sanctions are Non-deterrent”. *Scandinavian Journal of Economics*, vol. 108 (1), pp. 135–156.

van der Rijt, **Janine**, **Piet van den Bossche**, **Margje W. J. van de Wiel**, **Sven De Maeyer**, **Wim H. Gijselaers**, and **Mien S. R. Segers** (2013). “Asking for Help: A Relational Perspective on Help Seeking in the Workplace”. *Vocations and Learning*, vol. 6 (2), pp. 259–279.

van der Weele, **Joël J.**, **Julija Kulisa**, **Michael Kosfeld**, and **Guido Friebel** (2014). “Resisting Moral Wiggle Room: How Robust Is Reciprocal Behavior?” *American Economic Journal: Microeconomics*, vol. 6 (3), pp. 256–64.

Waldman, **Michael** (2012). “Theory and Evidence in Internal Labor Markets”. *The Handbook of Organizational Economics*. Princeton: Princeton University Press, pp. 520–571.

Yoon, **Jaewon**, **Grant Donnelly**, and **Ashley Williams** (2019). *It Doesn’t Hurt to Ask (for More Time): Employees Overestimate the Interpersonal Costs of Extension Requests*. HBS Working Paper 19-064.

Zepke, **Nick** and **Linda Leach** (2010). “Improving Student Engagement: Ten Proposals for Action”. *Active Learning in Higher Education*, vol. 11 (3), pp. 167–177.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Sofern ein Teil der Arbeit aus bereits veröffentlichten Papers besteht, habe ich dies ausdrücklich angegeben.

18.09.2019

Benjamin Tobias Häusinger